Investigating the Dynamic Risk Assessment for Offender Re-entry's (DRAOR) Ability to Operate as a Weighted Risk Prediction Instrument

by

Mackenzie Dunham

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

Master of Arts

in

Psychology

Carleton University
Ottawa, Canada

Abstract

Dynamic risk and protective factors refer to a collection of psychosocial variables that have been empirically linked to an increased or decreased likelihood of engaging in future criminal behaviour. Monitoring such factors is, therefore, a vital task in the post-incarceration community reintegration process. The current study examined whether weighting could augment the discrimination of the Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007, 2015, 2017), a promising case management instrument composed of dynamic risk and protective factors, in two samples of general justice involved individuals drawn from New Zealand (n = 3,648) and Iowa (n = 510). Two weighting approaches were investigated across subscales, outcomes, assessment periods, and samples. Although weighting did not significantly improve discrimination in either sample, the present research provides further support of the DRAOR's utility as a risk prediction and case management tool.

Keywords: DRAOR, dynamic risk assessment, parole, reintegration, case management

Acknowledgements

I would like to extend my sincerest gratitude to my supervisor, Dr. Ralph Serin, without whom this project would not have been possible. Your patience, support, and constructive advice have guided me throughout my graduate studies, and I am deeply appreciative of the numerous opportunities that you have provided. I consider myself to be extremely fortunate that you decided to take me on. I would also like to sincerely thank my committee member, Dr. Shelley Brown, for her insightful contributions that helped to develop this project, my external examiner, Dr. Katharine Kelly, for her assistance in refining this research, and my committee chair, Dr. Rachel Burns, for providing her time to oversee my thesis examination.

I am grateful to Dr. Nick Chadwick and Dr. Caleb Lloyd, who generously provided access to the two datasets that were used in the current research. This thesis would not have been possible without these invaluable resources. I am also very appreciative to Dr. Craig Leth-Steensen and Dr. Chris Lowenkamp for graciously lending their statistical expertise. Craig, your clear dedication, knowledge, and unwavering support significantly contributed to the success of this research. Special thanks are also extended to my previous mentors, Dr. Jessica Sutherland, Dr. John Turtle, and Dr. Julie Blais, who ensured that I had developed the foundational skills necessary to undertake this project.

My success throughout this process is also owed to my family, friends, and colleagues, whom I cannot thank enough. I withdrew great inspiration from your unconditional love, support, strength, encouragement, and compassion, and I consider myself to be truly indebted to each of you. Finally, I would like to acknowledge Nikki Kellermann-Lowe and Dr. Angela Smeth for granting me the workplace flexibility required to balance the demands of writing a thesis, and thank each of them for their personal investment in my professional development.

Table of Contents

List of Tables

List of Appendices

Investigating the Dynamic Risk Assessment for Offender Re-entry's (DRAOR) Ability to Operate as a Weighted Risk Prediction Instrument

Between 2012 and 2018, the percentage of Canadian federal justice involved individuals granted community supervision through day and full parole increased, and the total number of annual admissions to the Canadian federal jurisdiction decreased (Public Safety Canada, 2019). Encouragingly, revocation rates for day parole, full parole, and statutory release also decreased over the same period (Public Safety Canada, 2019),[1] indicating that community supervision outcomes are becoming increasingly successful prior to warrant expiration. Notably, however, revocation of community supervision constituted approximately one-third of annual admissions to the Canadian federal jurisdiction over the same period (Public Safety Canada, 2019), suggesting that there is still room for considerable improvement in the management and supervision of justice involved individuals who have been reintegrated back into the community. Comparably, although New Zealand has observed a gradually reducing rate of recidivism, room for improvement remains with 45.0% of justice involved New Zealanders facing reconviction and 29.8% facing reimprisonment within 12 months of their release (Ara Poutama Aotearoa, 2019). Similar figures have also been observed in the United States, where more than one-third of annual admissions and a half of prisoners overall are incarcerated as a result of revocation of a prior release (Klingele, 2013). Such a high revocation rate is unsurprising when considering the common challenges that prisoners face upon re-entry, including economic disadvantages, social exclusion, mental health problems, learning disabilities, substance misuse issues, political powerlessness, and a proneness to self-harm and suicide (Jewkes, 2007). In fact, as Travis and Petersilia (2001) explained, more systematic reintegration policies are needed to ensure that

[1] Where revocation is defined as an admission to federal custody following conditional release and prior to warrant expiry. Notably, this figure does not capture admissions to provincial/territorial jurisdictions.

justice involved individuals from disadvantaged communities are connected to meaningful community-based social structures and appropriately prepared for re-entry. Despite recent advances, these data from multiple jurisdictions suggests that the criminal justice system may be ineffectively managing those on community supervision.

Community supervision refers to the latter portion of a justice involved person's sentence that is served within the community under specific rules and restrictions. Generally, Canadians tend to be supportive of community supervision and believe that it results in reduced crime, increased public safety, and greater efficiency within the criminal justice system (Department of Justice, 2018). Although community supervision is often framed as an alternative to incarceration, it is important to note that it does not signal the end to a sentence (Klingele, 2013). Rather, justice involved individuals released on community supervision continue to be monitored within the community until at least their warrant expiry date or end of period of supervision. Community supervision may be granted in lieu of incarceration (i.e., probation), early through either day or full parole, or, less commonly, through statutory release (Parole Board of Canada, 2013).[2] Since dynamic risk assessments can be used to detect indicators of real change and predict the likelihood of success within the community, they play a crucial role in determining key decisions such as parole and supervision revocation.

The purpose of the present study was to investigate weighting strategies to attempt to maximize the predictive validity of a recently developed risk assessment and case management tool using two samples of general justice involved individuals under community supervision in

[2] In Canada, Statutory Release is granted to determinately sentenced justice involved individuals after they have served 2/3 of their sentence as long as they are not detained and not already on full parole (Public Safety Canada, 2019).

New Zealand and Iowa. The Dynamic Risk Assessment for Offender Re-entry (DRAOR; see Appendix A; Serin, 2007, 2015, 2017) is an emerging risk assessment and case management tool that was designed to assist in the assessment, management, and supervision of justice involved individuals. A growing body of literature has yielded promising results regarding the DRAOR, although further research is still needed. To date, the DRAOR has demonstrated predictive accuracy across various outcomes and with several different subsamples of justice involved individuals. The present research sought to expand the literature by reformulating the DRAOR to investigate whether it could effectively function as a weighted risk prediction instrument.

Within correctional research, risk assessment refers to the process of determining an individual's propensity to commit or recommit crime. Risk assessments can employ various methodologies, can be used to predict both general and crime-specific offending (e.g., violent offending, sexual offending), can be population-specific (e.g., justice involved youth, justice involved individuals with mental disorders), and can be environment-specific (e.g., institutional offending, community offending; Singh & Fazel, 2010). In addition to forecasting the likelihood of future offending, risk assessments can also be used to identify criminogenic needs (i.e., factors that are empirically linked to criminal involvement) and predict the severity of future offending. This is integral to the Risk-Need-Responsivity (RNR) model, which posits that treatment utility is maximized when correctional interventions are prioritized toward higher risk individuals, are designed to target criminogenic needs, and are delivered in a mode that complements the justice involved person's preferred learning style (e.g., cognitive, behavioural; Bonta & Andrews, 2017). Importantly, risk assessments provide invaluable information that assists with the management, rehabilitation, and reintegration of justice involved individuals.

Another factor that can be used to differentiate risk assessments is the type of risk that they measure. Risk factors refer to specific items or conditions that are associated with an increased susceptibility to engaging in criminal activity. Although some risk assessments incorporate both static and dynamic risk factors, others are limited to one type. Static risk refers to factors associated with criminal involvement that are invariant (e.g., age at onset of criminal behaviour, childhood maltreatment) and dynamic risk refers to changeable factors that are linked to criminal activity (e.g., substance misuse, unemployment; Bonta, 1996; McDermott et al., 2008; Zamble & Quinsey, 1997). Interestingly, some researchers choose to treat some dynamic risk factors as static (e.g., the "central eight" criminogenic risk factors; Bonta & Andrews, 2017) because they tend to remain relatively stable across time. In part, this reflects how the construct is measured, rather than being theoretically changeable (Serin et al., 2016).

Risk assessments also differ in their approach; specifically, assessments can be based on either actuarial prediction or clinical judgement, which can be further broken down into unstructured clinical judgement and structured professional judgement (Hanson, 2009). Actuarial prediction prevents human influence by basing decisions solely upon empirically established relationships between variable clusters and the event of interest (Dawes et al., 1989). Unstructured clinical judgement is a contemporary label that closely resembles earlier conceptualizations of clinical prediction, which Meehl (1954) described as a decision based primarily on the assessor's own experiences, opinions, and intuitions. Structured professional judgement may be best understood as a hybrid approach, as it relies on predefined risk factors with explicit coding procedures that allow for slight modification based on the assessor's interpretation of available evidence. Importantly, decisions made in clinical settings should not be automatically labelled as clinical judgement, because many clinicians routinely practice

actuarial prediction (Dawes et al., 1989). While previous research has found that structured professional judgement assessments perform with moderate accuracy (Wilson et al., 2013; Singh et al., 2011), meta-analytic data indicates that actuarial assessments have an average 13% increase in accuracy compared to unstructured clinical judgement (Ægisdóttir et al., 2006).

The present study focuses on reformulating an emerging structured professional judgement measure, the DRAOR, with the goal of deriving weighted items to examine its utility as a validated actuarial scale.

Scale Validation

The ultimate goal of risk assessment is to accurately predict an individual's recidivism outcome. Although there are several aspects that contribute to the accuracy of a statistical model, discussion of the discrimination and calibration components of a model often prevail (Cook, 2008; Helmus & Babchishin, 2017). Specific to risk assessments, the discrimination component of a model refers to its ability to separate non-recidivists from recidivists. The calibration component, on the other hand, refers to a risk assessment's ability to correctly estimate the probability of a future event, such as the magnitude of recidivism risk (e.g., low, moderate, or high). Once a scale appears to have sufficient discrimination and calibration, a necessary part of the scale validation process is determining whether the scale has construct validity. Construct validity refers to the degree to which a scale is capable of measuring the target concept, with the goal of having a high degree of correspondence between a scale's predicted outcome and the actual observed outcome (American Psychological Association [APA], 2021). Establishing construct validity is integral to scale development because it helps to ensure that a study's conclusions have merit and can be trusted. This is especially true in psychological research, which deals with unobservable constructs, known as latent variables, that can typically only be

inferred (Flake et al., 2017). For example, correctional researchers often rely on risk assessment scores as a proxy for the unobservable construct of recidivism risk; notably, however, these risk scores cannot be used to inform research or case planning unless we are confident that they accurately forecast the recidivism outcome. Flake and colleagues (2017) outline the scale validation process used in social research, summarized below.

Construct validation is broken down into three sequential phases: (1) substantive, (2) structural, and (3) external. The substantive phase is a reconceptualization of Loevinger's (1957) concept of content validity, comprising the theoretical underpinnings of a scale to ensure that all facets of a construct are adequately measured. Typically, a robust literature review is conducted to ensure that a new scale is needed and to operationally define an overly inclusive pool of potential items for consideration (Gehlbach & Brinkworth, 2011). Another primary goal of the substantive phase is to identify the depth and breadth of a construct (Gehlbach & Brinkworth, 2011), which will assist with defining the scope of the scale and subsequently help to determine which items should be included. Although less common, other data collection methods that can be employed in the substantive phase include expert reviews and focus groups (Salerno et al., 2012). Overall, the substantive phase seeks to outline all the necessary information needed to reasonably measure the construct.

The structural phase of construct validation examines the psychometric properties of a scale using quantitative analyses, including measures of internal consistency, reliability, measurement invariance, and factor analysis. Measures of internal consistency (e.g., Cronbach's alpha, item-total correlations) evaluate the homogeneity of scale items, indicating whether they capture various aspects of the same construct (APA, 2021). In discussing the predictive accuracy of risk assessments, Helmus and Babchishin (2017) contend that predictive accuracy is

maximized when the number of variables introduced into a model is minimized, which ultimately increases efficiency and decreases internal consistency. Measures of reliability indicate the degree of consistency between ratings that occur across independent evaluators of the same individual (i.e., inter-rater reliability) or from one time to another (i.e., test-retest reliability; APA, 2021). Measurement invariance testing ensures that the target construct is measured in the same (or a sufficiently similar) way across several different subgroupings of individuals (APA, 2021). Factor analysis refers to a procedure in which a large group of variables is reduced into a simpler cluster of variables that effectively represent the target construct (Field, 2009). As explained by Helmus and Babchishin (2017), reliably examining the factor structure of risk scales can be difficult or ineffective if too few variables were introduced into the original model. Additionally, traditional elements of scale validation, such as analyses of factor structure, may not provide meaningful insight for risk assessments due to their criterion-referenced nature (e.g., the intention to predict a specific outcome; Helmus & Babchishin, 2017). Importantly, the pilot testing that occurs in this phase must include all focal items (i.e., those that are believed to be necessary to include in the scale), as well as items that can used to establish scale validity, and be followed by analyses that seek to eliminate problematic or unnecessary items (Gehlbach & Brinkworth, 2011).

The external phase of construct validation involves gathering evidence that helps to better understand the target construct by placing it within a larger nomothetical network, through which observations can be discussed in relation to each other, the target construct, and theoretically similar constructs (Cronbach & Meehl, 1955). The external phase includes evaluations of convergent and divergent validity (also known as congruent and discriminant validity, respectively), criterion validity, and a detection of differences between groups that have been

previously demonstrated to differ on the construct. Convergent validity refers to the extent to which the scale aligns with theoretically similar measures, whereas divergent validity refers to the extent to which the scale deviates from theoretically dissimilar measures (APA, 2021). Divided into three subtypes, criterion validity is a measure of how well a scale can use current scores to predict past (i.e., retrospective validity), current (i.e., concurrent validity), and future (i.e., predictive validity) outcomes related to the target construct (APA, 2021). In the context of risk assessment, sufficient criterion validity would require the ability to predict whether an individual (1) has a previous criminal record, (2) is currently breaking the law, and (3) will commit crimes in the future (APA, 2021).

As summarized by Gehlbach and Brinkworth (2011), validity is not an end state, and the idea of a fully validated scale is a misnomer. The results of the aforementioned process will necessarily fluctuate depending upon the different participants, contexts, and applications that are involved across validation studies. As such, it is essential that a new scale consistently demonstrates reliability and validity over several investigations that differ in scale application, context, and participant selection, even in the presence of a growing body of research demonstrating promising results. Notably, the scale validation process may include different measures and elements, depending on the specific goals, focus, and applications of the scale (e.g., risk prediction versus case management; Helmus & Babchishin, 2017; Perley-Robertson, 2018). Nonetheless, according to Gehlbach and Brinkworth (2011), scale developers have reached a broad consensus that a scale can be considered an adequate measure of a construct if it (1) minimizes respondent error, and (2) demonstrates sufficient reliability. With respect to the DRAOR, the constructs in question would be stable risk, acute risk, and protective factors. Therefore, from Gehlbach and Brinkworth's (2011) perspective, the DRAOR could be

considered to be an adequate measure if it demonstrates sufficient reliability in assessing stable risk, acute risk, and protective factors with minimal respondent error. In this context, successful supervisees would be expected to be lower scoring in both risk domains and higher scoring in the protective domain (resulting in lower DRAOR Total scores), and unsuccessful supervisees (i.e., recidivists and technical violators) would be expected to be higher scoring in both risk domains and lower scoring in the protective domain (resulting in higher DRAOR Total scores). Meanwhile, Helmus and Babchishin (2017) suggest that sufficient predictive validity alone can singlehandedly legitimize the application of a risk assessment tool.

Configuration and Implementation of the DRAOR

The DRAOR is a 19-item case management instrument that was designed to assist community supervision officers (CSOs; e.g., parole and probation officers) in the assessment and reassessment of a justice involved person's risk state at various points of supervision. Although many CSOs already consider some of the DRAOR's items during their interactions with justice involved individuals, considerably less attention is typically given to protective items. Ideally, the introduction of a structured tool that includes both risk and protective items, such as the DRAOR, will encourage CSOs to make more fulsome assessments. The DRAOR should be used at each point of substantive contact during supervision to assess, manage, and monitor the current experiences of justice involved individuals over time. If used this way, the DRAOR can putatively yield valuable information that could assist CSOs in developing and updating individualized risk management strategies that are sensitive to the justice involved individual's current state. Additionally, frequent use of the DRAOR can show real-time changes of risk (or lack thereof), which may be useful in monitoring rehabilitative progress over time and assist in projecting future risk (Lloyd et al., 2020a).

The DRAOR is organized into three domains: Stable, Acute, and Protective. These domains organize items based on their relationship with desistance (i.e., Protective items) or reoffending (i.e., Stable and Acute items). While Stable and Acute items are both associated with a risk of reoffending, they are distinguished by their degree of stability. Since the DRAOR follows a structured professional judgement approach, the assessor must follow the scale's guidelines and use all available relevant evidence to determine scores for each item. Each item is scored on a three-point scale (0-2). For Stable and Acute items, a score of 0 indicates that the client does not currently present signs of the associated risk factor, a score of 1 indicates either a slight problem or uncertainty due to mixed evidence, and a score of 2 indicates a definite problem. For Protective items, a score of 0 indicates that the client is currently absent of the associated protective factor, a score of 1 indicates either slight protection or uncertainty due to mixed evidence, and a score of 2 indicates a definite asset. Item scores can be used to identify target areas that may require treatment intervention. Total scores on the DRAOR can be derived by adding the scores for Stable and Acute items and subtracting the scores for Protective items, yielding a Total score range from -12 to 26. Lower Total scores indicate that the individual has fewer risk factors and more protective factors, whereas higher Total scores indicate that the individual has more risk factors and fewer protective factors. Therefore, lower Total scores are theoretically expected to be associated with more successful correctional outcomes and vice versa.

Understanding the relationship between DRAOR scores and correctional outcomes (e.g., technical violations, recidivism) is necessary to ensure that the scale is effectively applied. For example, since each item was included in the DRAOR due to its empirical relationship with either reoffending or desistance, CSOs should be able to use DRAOR scores to detect real-time

changes in risk state. To facilitate this, Serin and Chadwick (2017) developed cut-off scores through the visual inspection DRAOR data from the Iowa Department of Corrections. Although the proposed cut-offs have yet to be validated, they provide simple guidelines to assist CSOs in appropriately adjusting supervision intensity. These cut-offs utilize DRAOR Total scores to categorize justice involved individuals into one of four risk levels: (1) Low, (2) Moderate, (3) Moderate/High, and (4) High (see Appendix B). Serin and Chadwick's (2017) proposed cut-offs can then be used to adjust the client's initial supervision level that was based on static risk at the beginning of supervision; specifically, a low DRAOR Total score encourages modifying supervision down by one level, a moderate/high or high DRAOR Total score encourages modifying supervision up by one level, and a moderate DRAOR Total score encourages maintaining the current supervision level.

Stable Items

Stable dynamic risk refers to variables that are changeable but usually endure for months or years (Public Safety Canada, 2007). The DRAOR includes six Stable items: (1) Peer Associations, (2) Attitudes Toward Authorities, (3) Impulse Control, (4) Problem-Solving, (5) Sense of Entitlement, and (6) Attachment with Others.

Acute Items

Acute dynamic risk refers to variables that are changeable within minutes, hours, days, or weeks (Public Safety Canada, 2007). The DRAOR includes seven Acute items: (1) Substance Abuse, (2) Anger/Hostility, (3) Opportunity/Access to Victims, (4) Negative Mood, (5) Employment, (6) Interpersonal Relationships, and (7) Living Situation.

Protective Items

Protective factors refer to assets or skills that a justice involved person possesses that may reduce their risk of reoffending (Ullrich & Coid, 2011). Protective factors can be either internal (e.g., responsiveness to advice) or external (e.g., social support) to the individual. Importantly, protective factors do not simply denote a lack of risk; rather, protective factors are specific characteristics that encourage a withdrawal from criminal behaviour in favour of prosocial participation in society. The DRAOR includes six Protective items: (1) Responsive to Advice, (2) Prosocial Identity, (3) Realistic High Expectations, (4) Cost/Benefits (Supportive of Staying Crime-Free), (5) Social Support, and (6) Social Control. More fulsome descriptions of each DRAOR item are available in Appendix C.

Initial Findings with the DRAOR

The DRAOR has been implemented and tested in several international jurisdictions, giving rise to a growing body of psychometric data. Previous research has found the DRAOR to be predictive of recidivism among general justice involved individuals (Chadwick, 2014; Hanby, 2013; Lloyd, 2015; Tamatea & Wilson, 2009), high-risk justice involved individuals (Yesberg & Polaschek, 2015), justice involved individuals who have committed a sexual offence (Averill, 2016; Smeth, 2013),[3] justice involved youth (Ferguson, 2015), justice involved women (Scanlan, 2015; Scanlan et al., 2020; Yesberg et al., 2015), justice involved partner-violent men (Perley-Robertson et al., 2020),[4] justice involved Māori (Hanby, 2013), and justice involved individuals with or without a mental disorder (Wardrop, 2020). The subsequent section provides a brief

[3] Notably, neither study found the DRAOR to significantly predict sexual recidivism, although it was predictive of parole violations (Smeth, 2013), violent recidivism (Averill, 2016), and general recidivism (Averill, 2016) using a sample of justice involved individuals who committed a sexual offence.

[4] The DRAOR did not significantly predict intimate partner violence recidivism, but it was predictive of general recidivism, violent recidivism, and technical violations in a sample of partner-violent men on community supervision (Perley-Robertson et al., 2020). This sample overlapped with the sample used by Perley-Robertson (2018).

summary of these results, with particular focus on the psychometric properties and predictive validity of the DRAOR.

Item Analyses

Item analysis refers to a set of procedures that aims to evaluate the merit of each individual item that comprises a scale (APA, 2021). Item analyses can be conducted during scale development to help determine items for inclusion or during scale validation to gauge the composition of an emerging scale (APA, 2021). Since the quality of included items determines the reliability and validity of a scale, conducting item-level analyses can yield considerable insight into scale optimization. There are several different approaches to conducting an item analysis, although the most common approach in psychological research is factor analysis (Ellis, 2017). Factor analysis refers to a broad range of procedures designed to reduce the set of interrelations among manifest (i.e., directly observable) variables to a smaller set of latent (i.e., unobservable) variables (APA, 2021). Notably, the DRAOR domains were not derived from factor analysis (Perley-Robertson, 2018), but subsequent DRAOR validation studies have since conducted analyses to assess the DRAOR's factor structure.

To date, multiple studies have conducted factor analyses to examine the factor structure of the DRAOR in an exploratory fashion. Using a representative sample (N = 3,498) of New Zealand parolees, Hanby (2013) conducted an exploratory factor analysis on over 97,000 DRAOR assessments. Overall, it was suggested that three factors (i.e., Mostly Stable, Mostly Acute, and Protective) should be retained. Collectively, these three factors retained 17 of the DRAOR's original 19 items, dropping the Employment and Interpersonal Relationships items. This model explained 38.2% of the total variance. Additionally, a confirmatory factor analysis conducted by Hanby (2013) supported that the original three-factor structure of the DRAOR

adequately fit her data as well. Notably, however, Yesberg and Polaschek (2015) also conducted a confirmatory factor analysis on the original three-factor structure of the DRAOR using a sample of 299 high-risk male parolees from New Zealand, which revealed a poor fitting model. Next, Yesberg and Polaschek (2015) conducted a principal components analysis, which yielded a four-factor model. This model, comprised of Protective, Stable, Internal Acute, and External Acute domains, retained all 19 of the original DRAOR items and explained 51.3% of the total variance. Using a representative sample (N = 391) of all parolees and probationers in the 5th District of Iowa, Chadwick's (2014) exploratory factor analysis yielded a two-factor model. This model, comprised of Risk and Protective domains, retained all 19 of the original DRAOR items. Following Chadwick's research, Wardrop (2020) conducted a confirmatory factor analysis on this two-factor model using a sample of 961 adult male and female justice involved individuals from the state of Iowa. The model was found to have acceptable-to-good fit and the majority of the DRAOR items loaded strongly onto their respective factors.

Evidence of DRAOR Reliability

Research has shown that DRAOR scores are normally distributed and demonstrate change over the course of supervision (Chadwick, 2020; Lloyd et al., 2020a; Tamatea & Wilson, 2009). However, previous investigations into the DRAOR's internal consistency, a common indicator of scale reliability that measures whether a series of items have been appropriately grouped based on their interrelatedness, have yielded consistently mixed results. For example, Hanby (2013) found acceptable internal consistency for the Stable (Cronbach's $\alpha = .81$) and Protective (Cronbach's $\alpha = .84$) domains in a sample of 3,498 New Zealand parolees, [5] although

[5] Cronbach's alpha values of .70 and above indicate acceptable internal consistency when used in the early stages of research (e.g., scale development; Nunnally, 1978). However, Nunnally (1978) recommends a more conservative cut-off value of .80 for use in basic research.

the Acute domain yielded considerably lower internal consistency (Cronbach's α = .62).[6] Likewise, Chadwick (2014) observed nearly identical results in a sample of 391 probationers and parolees in the state of Iowa for the Stable (Cronbach's α = .81), Acute (Cronbach's α = .62), and Protective domains (Cronbach's α = .86). More recently, similar trends of internal consistency have been observed among general (Wardrop, 2020)[7] and intimate partner violent (Perley-Robertson, 2018) parolees in the state of Iowa as well.

Another common indicator of scale reliability is item-total correlations. Item-total correlations are a measure that determines whether a particular item score is sufficiently correlated with the subscale's total score. In the above-mentioned research, item-total correlations were generally acceptable.[8] The Stable subscale has been routinely observed to have acceptable item-total correlations, with Hanby (2013) reporting a range of .47 to .63, Chadwick (2014) reporting a range of .43 to .68, Perley-Robertson (2018) reporting a range of .49 to .69, and Wardrop (2020)[9] reporting a range of .68 to .81. Likewise, the Protective subscale has been routinely observed to have acceptable item-total correlations as well, with Hanby (2013) reporting a range of .52 to .57, Chadwick (2014) reporting a range of .57 to .68, Perley-Robertson (2018) reporting a range of .61 to .70, and Wardrop (2020) reporting a range of .70 to .87. However, markedly poorer item-total correlations have been observed for the Acute subscale, with Hanby (2013) reporting a range of .20 to .46, Chadwick (2014) reporting a range of .28 to .42, Perley-Robertson (2018) reporting a range of .28 to .52, and Wardrop (2020) reporting a range .47 to .65. Item-total correlations for the Acute subscale that were below the

---

[6] Due their time-dependent nature, acute dynamic risk factors may not optimally contribute to internal consistency. Additionally, recall that Helmus and Babchishin (2017) contend that high internal consistency may impinge on the accuracy of risk prediction instruments, so lower internal consistency on the Acute domain may not be of concern.
[7] Notably, Wardrop (2020) observed a degradation in assessment fidelity and decreasing staff engagement over time.
[8] Item-total correlations of .30 and above are considered acceptable (Field, 2009).
[9] Item-total correlations were calculated separately for mentally disordered and non-disordered justice involved individuals at baseline and over time. These ranges are inclusive of calculations for both subsamples at baseline.

acceptable level included Interpersonal Relationships (Hanby, 2013), Employment (Hanby, 2013; Perley-Robertson, 2018), and Substance Abuse (Chadwick, 2014; Perley-Robertson, 2018).

Little is presently known regarding the inter-rater reliability of DRAOR assessments. Inter-rater reliability is a measure that determines the level of consistency between ratings done on the same cases by different assessors. Although researchers have yet to investigate the inter-rater reliability of the DRAOR, the Structured Dynamic Assessment Case-management 21-item (SDAC-21; Serin & Wilson, 2013, 2017),[10] an institutional assessment tool based on the DRAOR, has demonstrated fair to excellent intra-class correlations (ICCs ranged from .54 to .96; Smeth, 2019).[11] This research may lend some support to the DRAOR, although future research is needed to provide direct evidence.

Evidence of DRAOR Validity

Previous research has indicated that the DRAOR's subscales are highly inter-correlated.[12] For example, the DRAOR Total score has been strongly correlated with the Stable (r = .87 to .92), Acute (r = .81 to .87), and Protective subscales (r = -.82 to -.89; Averill, 2016; Ferguson, 2015; Muirhead, 2016; Perley-Robertson, 2018; Smeth, 2013). Additionally, the DRAOR Stable and Acute subscales have been consistently found to be moderately to strongly positively correlated (r = .46 to .73), indicating that the two subscales measure similar constructs (i.e., criminogenic risk; Averill, 2016; Chadwick, 2014; Chadwick, 2020; Ferguson, 2015; Muirhead, 2016; Perley-Robertson, 2018; Smeth, 2013; Tamatea & Wilson, 2009; Wardrop, 2020).

---

[10] The SDAC-21 and the DRAOR share 16 overlapping items, including all six DRAOR Stable items, all six DRAOR Protective items, and four DRAOR Acute items.

[11] Cicchetti (1994) suggests that ICCs below .40 indicate poor inter-rater reliability, ICCs between .40 and .59 indicate fair inter-rater reliability, ICCs between .60 and .74 indicate good inter-rater reliability, and ICCs above .75 indicate excellent inter-rater reliability.

[12] Cohen (1992) suggests that correlations of .10, .30, and .50 represent small, moderate, and large effect sizes, respectively.

Conversely, researchers have consistently reported negative associations between the Protective subscale and the Stable (r = -.40 to -.73) and Acute subscales (r = -.27 to -.63), suggesting that the Protective subscale is composed of factors that operate differently than dynamic risk factors do (Averill, 2016; Chadwick, 2014; Chadwick, 2020; Ferguson, 2015; Muirhead, 2016; Perley-Robertson, 2018; Smeth, 2013; Tamatea & Wilson, 2009; Wardrop, 2020).

Evidence of DRAOR Construct Validity. Previous researchers have also determined that the DRAOR has sufficient construct validity. Construct validity can be further subdivided into convergent and divergent validity (APA, 2021). Convergent validity tests whether two theoretically linked measures are actually related and divergent validity tests whether two theoretically unrelated measures are actually unrelated. Together, acceptable levels of convergent and divergent validity work to establish construct validity, which indicates the degree to which a scale measures what it is designed to measure.

In an initial validation study using a sample of 59 New Zealand probationers, Tamatea and Wilson (2009) compared the DRAOR to the Risk of re-Conviction X Risk of re-Imprisonment model (ROC*ROI; Bakker et al., 1999), a validated measure of static risk that has been implemented in New Zealand. Results indicated that the DRAOR's Protective subscale was negatively correlated with the ROC*ROI (r = -.33), indicating that having a greater number of protective factors is associated with lower levels of static risk. The ROC*ROI was positively associated with the Stable (r = .22) and Acute (r = .03) subscales of the DRAOR, although neither of these associations were significant. In a more robust study of 3,498 justice involved individuals, Hanby (2013) found similar trends between the DRAOR and ROC*ROI. The study used scores from both baseline (i.e., the first assessment) and proximal assessments (i.e., the most recent assessment prior to either recidivism or the study end date), yielding significant

correlations in the expected directions for the Total (*r* = .32, .37, respectively), Stable (*r* = .31, .35, respectively), Acute (*r* = .24, .30, respectively), and Protective subscales (*r* = -.25, -.31, respectively). Similarly, Averill (2016) reported an association between the DRAOR and the ROC*ROI in a sample of 851 male parolees who had previously committed a sexual offence in New Zealand, with the ROC*ROI being significantly correlated with the DRAOR Total (*r* = .38) and the Stable (*r* = .37), Acute (*r* = .40), and Protective domains (*r* = -.25).

The DRAOR has also demonstrated convergent and divergent validity with several other risk instruments, including the Static-99R (an actuarial risk scale used to assess sexual recidivism risk, Phenix et al., 2016), the Violence Risk Scale (VRS; a static and dynamic risk scale used to assess violence and general risk; Wong & Gordon, 1999-2003), the Release Proposal Feasibility Assessment-Revised (RPFA-R; a dynamic risk scale used to assess an inmate's preparedness for release; Wilson, 2011), the Automated Sexual Recidivism Scale (ASRS; a static risk scale used to assess sexual recidivism risk; Skelton et al., 2006), and the Iowa Violence and Victimization Instrument (IVVI; an institutional static risk scale developed by the Iowa Department of Corrections to assist in assessing violent recidivism risk). Using a sample of 193 justice involved Iowan men, Smeth (2013) found the Static-99R to be significantly correlated with the DRAOR Total (*r* = .26), as well as the Stable (*r* = .23), Acute (*r* = .17), and Protective subscales (*r* = -.24). Using a sample of 299 high-risk justice involved men on community supervision in New Zealand, Yesberg and Polaschek (2015) found the VRS Dynamic domain to be correlated with the DRAOR Total (*r* = .25), as well as the Stable (*r* = .18), Acute (*r* = .22), and Protective subscales (*r* = -.20). The researchers also reported significant correlations between the VRS Static domain and the DRAOR Total (*r* = .16) and the Protective subscale (*r* = -.19). Additionally, Yesberg and Polaschek (2015) found significant correlations between the RPFA-R

and the DRAOR Total (*r* = .30) and the Stable (*r* = .20), Acute (*r* = .34) and Protective domains (*r* = -.17). Using a sample of 851 community-supervised Iowan men who had offended sexually, Averill (2016) found the ASRS to be correlated with the DRAOR Total (*r* = .47), as well as the Stable (*r* = .45), Acute (*r* = .45), and Protective subscales (*r* = -.37). Finally, Wardrop (2020) used a sample of 961 justice involved adult males and females from the state of Iowa to compare DRAOR Total and IVVI Victimization scores at 1-2 months (Time 1), 3-4 months (Time 2), and 5-6 months (Time 3) post-release. Wardrop (2020) found the scores to be moderately (*r* = .36; Time 1) to strongly correlated (*r* = .79, .84; Times 2 and 3, respectively).

Overall, the DRAOR and its subscales have demonstrated small to moderate significant correlations with several other validated risk skills, including the ROC*ROI, Static-99R, RPFA-R, ASRS, IVVI, and VRS (although the VRS Static domain did not significantly correlate with the DRAOR Stable and Acute subscales). The Stable subscale identifies several criminogenic needs, making it similar to traditional risk-need scales. Although the Acute subscale also identifies risk factors associated with community supervision failure, these items are, theoretically, more transient. The Protective subscale identifies factors that are associated with successful community supervision, and the pattern of negative correlations between this subscale and other measures of dynamic risk indicates that its items are in fact dissimilar from dynamic risk items.

Evidence of DRAOR Predictive Validity. Predictive validity refers to the degree to which a score on an assessment scale forecasts a subsequent outcome. As summarized by Table 1, several studies have demonstrated that DRAOR assessments are predictive of various types of recidivism. Hanby (2013), for example, found the DRAOR to be significantly predictive of general recidivism in a sample of 3,372 general justice involved adult males and females in New

Zealand. Results indicated that DRAOR subscale and Total scores were predictive of any criminal reconviction (AUC values ranged from .62 to .67, small to moderate effects)[13] as well as any reconviction (AUC values ranged from .66 to .72, moderate to large effects).[14] Hanby (2013) also found that the DRAOR added incremental validity to the ROC*ROI's prediction of any reconviction and any criminal reconviction, signifying that, when administered together, the DRAOR significantly improves the ROC*ROI's predictive accuracy.[15] More recently, Serin and colleagues (2020) reported the DRAOR to have moderate predictive utility in a representative sample of 563 probationers and parolees in the state of Iowa. Overall, the DRAOR subscale and Total scores significantly predicted technical violations (AUCs ranging from .66 to .71, moderate to large effects), serious parole violations (AUCs ranging from .66 to .72, moderate to large effects), and new crimes (AUCs ranging from .67 to .70, moderate effects).

Additional research by Yesberg and colleagues (2015) employed a subset of the dataset used by Hanby (2013) to evaluate differences in the DRAOR's predictive accuracy between sexes ($N = 266$; 133 male and 133 female parolees).[16] Results indicated that the DRAOR Total score significantly predicted time to recidivism for both men and women. Interestingly, however, results also indicated that the DRAOR may be more robustly predictive for justice involved women in three key areas that were significant for women but not men: (1) collectively, the Stable, Acute, and Protective subscales were predictive of time to recidivism, (2) the Acute subscale made a unique contribution to the prediction of time to recidivism, and (3) the Total

[13] There are no clear AUC cut-offs. However, Rice and Harris (2005) recommend using .56, .64, and .71 as cut-offs for small, moderate, and large effect sizes (respectively) in forensic psychology and psychiatry. These values were chosen because they correspond with commonly used cut-off values for two other longstanding measures of effect size (i.e., Cohen's d and the Point-Biserial Correlation Coefficient).
[14] Any reconviction was defined as any reconviction resulting from either a criminal or administrative offence.
[15] Averill (2016) also reported that the DRAOR showed incremental validity beyond the ROC*ROI in the prediction of violent and general recidivism, once again demonstrating the DRAOR's ability to improve the predictive accuracy of this static risk assessment.
[16] Notably, males had markedly lower recidivism base rates compared to females (16% vs 26%, respectively).

score demonstrated incremental predictive validity after accounting for static risk. The researchers theorized that the large proportion of breach of parole violations in their sample could help to explain why the DRAOR appeared to be more predictive for justice involved women. Notably, this theory was primarily based on earlier research that demonstrated that DRAOR Total scores were not predictive of breaches of parole among high-risk justice involved men (Yesberg & Polaschek, 2015) or general justice involved individuals (Tamatea & Wilson, 2009).[17] However, more recent research has consistently demonstrated DRAOR Total scores to be significantly predictive of serious parole violations (Serin et al., 2020) and technical violations (Chadwick, 2014; Ferguson, 2015; Perley-Robertson et al., 2020; Serin et al., 2020; Smeth 2013; Wardrop, 2020) using a variety of samples of justice involved individuals. Therefore, subsequent research has yet to uncover support for this theory, and it remains unclear why the DRAOR demonstrated greater predictive utility for justice involved women in this early validation study.

Similar findings that the DRAOR may function better for women were also reported by Scanlan (2015), who used an overlapping sample to determine that both initial (AUC = .60) and proximal (AUC = .64) DRAOR scores were predictive of breach and criminal reconvictions in women, whereas only proximal (AUC = .63) DRAOR scores were predictive of reconvictions in men.[18] Most recently, Scanlan and colleagues (2020) employed a matched sample of justice involved individuals serving community supervision sentences (n = 350) to further investigate differences in the DRAOR's predictive validity across men (n = 175) and women (n = 175).[19]

---

[17] That said, both of these studies found the Protective subscale to significantly predict breaches of parole.
[18] Scanlan's (2015) sample partially overlapped with the samples used by Yesberg and colleagues (2015) and Hanby (2013).
[19] Participant matching controlled for differences in ethnicity, index offence, age, sentence length, ROC*ROI score, and number of previous convictions, violent convictions, and imprisonments.

Overall, the DRAOR Total scores were equally predictive for men (AUC = .63) and women (AUC = .64), although sex differences emerged in the prediction of reconvictions using the Stable (AUC_{men} = .59; AUC_{women} = .63), Acute (AUC_{men} = .62; AUC_{women} = .58), and Protective subscales (AUC_{men} = .59; AUC_{women} = .64; Scanlan et al., 2020). While no subscale uniquely predicted reconvictions for women, the Acute subscale independently predicted reconvictions for men (Scanlan et al., 2020). Notably, this finding directly contrasts the previous finding that the Acute subscale is more predictive for women (Yesberg et al., 2015), suggesting that further research into the DRAOR's predictive validity across sexes is warranted.

The DRAOR has also demonstrated predictive validity among justice involved youth in two studies. First, Ferguson (2015) used a matched sample of justice involved youth (aged 17-19; *n* = 455) and adult (aged 20-60; *n* = 549) males on community supervision in New Zealand to examine the predictive validity of initial and proximal DRAOR assessments.[20] For youth, results indicated that initial DRAOR Total scores significantly predicted criminal reconvictions (AUC = .64, a moderate effect), but not reconvictions due to breaches of conditions. None of the initial DRAOR subscores demonstrated significant predictive accuracy for either outcome. Proximal DRAOR subscale and Total scores demonstrated significant predictive accuracy for both criminal (AUC values ranged from .65 to .74, moderate to large effects) and breach reconvictions (AUC values ranged from .65 to .72, moderate to large effects). For adults, results indicated that initial DRAOR Total and Acute scores significantly predicted criminal (AUCs = .62 and .63, respectively, small effects) and breach reconvictions (AUCs = .66 and .71, respectively, moderate and large effects, respectively). Neither of the initial DRAOR Stable and Protective subscores demonstrated significant predictive accuracy. Proximal DRAOR subscale

[20] Participant matching controlled for differences in ethnicity, index offence, sentence length, static risk, and number of prior convictions, violent convictions, and imprisonments.

and Total scores significantly predicted both criminal (AUC values ranged from .63 to .65, small to moderate effects) and breach reconvictions (AUC values ranged from .68 to .77, moderate to large effects). Second, Muirhead (2016) used a sample of justice involved male (*n* = 455) and female (*n* = 92) youths (aged 17-19) on community supervision in New Zealand to examine the predictive validity of initial and proximal DRAOR assessments.[21] Initial DRAOR subscale and Total scores significantly predicted both violent and any criminal reconvictions (AUCs for both outcomes exhibited moderate effects, ranging from .59 to .63). Proximal DRAOR subscale and Total scores significantly predicted both violent (AUC values ranged from .61 to .68, small to moderate effects) and any criminal reconvictions (AUC values ranged from .68 to .73, moderate to large effects).

Recently, Wardrop (2020) investigated the DRAOR's predictive accuracy among a sample of justice involved Iowan men and women with (*n* = 478) and without (*n* = 483) a diagnosed mental disorder. Although the DRAOR was able to discriminate non-disordered recidivists and non-recidivists with moderate accuracy (AUC values ranged from .63 to .69, small to moderate effects), Wardrop reported less favourable observations among mentally disordered justice involved individuals. Specifically, the DRAOR assessed mentally disordered justice involved individuals as having significantly lower levels of Protective factors and significantly higher levels of dynamic risk, ultimately suggesting that, when administered to mentally disordered justice involved individuals, the DRAOR may under-classify lower-scoring individuals and over-classify higher-scoring individuals. Overall, Wardrop advised that the DRAOR should be used with caution when assessing justice involved individuals with diagnosed mental disorders.

[21] Many of the participants of this study overlapped with those from Ferguson (2015).

Most recently, Serin and colleagues (2020) conducted a quasi-field study (see Edens & Boccaccini, 2017) to examine the DRAOR's utility in predicting recidivism. Using a sample of 510 justice involved men who were serving community supervision orders in Iowa, Serin and colleagues (2020) found that each of the DRAOR's domains were significant univariate predictors of both technical violations and new charges. Further, the authors reported that the DRAOR uniquely contributed to the prediction of recidivism beyond the Iowa Risk Assessment (IRA; Iowa Department of Corrections, 2003), a static risk scale used to determine the appropriate level of supervision for a justice involved individual. Specifically, the DRAOR Protective and Acute domains added incremental validity to the IRA in the prediction of technical violations and the DRAOR Protective and Stable domains added incremental validity to the IRA in the prediction of new charges (Serin et al., 2020).

Overall, previous research has demonstrated that the DRAOR can generally be applied to diverse populations of justice involved individuals, regardless of their sex, age, and ethnicity. The DRAOR has demonstrated the ability to operate as a risk prediction instrument and has added incremental predictive validity to multiple measures of static risk. Notably, however, the DRAOR does not appear to be an effective measure for use on justice involved individuals with diagnosed mental disorders, especially among lower- and higher-scoring individuals. Small to large effect sizes have been observed for general recidivism, technical violations, and any recidivism, and small to moderate effect sizes have been observed for violent recidivism. Mixed results have been reported regarding the DRAOR's utility in predicting sexual and intimate partner violent recidivism, warranting further investigation into these areas. In summary, although further research is still needed, previous research corroborates the continued use of the DRAOR as a case management tool.

Table 1

A Summary of Previous Research on the Predictive Validity of the DRAOR and its Subscales for Various Recidivism Outcomes

			Stable		Acute		Protective		Total	
Source	*N*	Follow-Up[a]	*r*	AUC [95% CI]	*r*	AUC [95% CI]	*r*	AUC [95% CI]	*r*	AUC [95% CI]
Any recidivism (including supervision violations)										
Hanby (2013)[b]	1,775	2 years	-	.65 [.62, .67]	-	.70 [.68, .72]	-	.65 [.62, .67]	-	.69 [.67, .72]
Hanby (2013)[c]	3,372	2 years	-	.66 [.65, .68]	-	.72 [.70, .74]	-	.67 [.65, .68]	-	.71 [.69, .73]
Chadwick (2014)[d]	391	*M* = 250 days	-	.62 [.56, .67]	-	.59 [.54, .65]	-	.58 [.52, .66]	-	.62 [.56, .68]
Yesberg & Polaschek (2015)[e]	299	6 months	-	.59* [-, -]	-	.60* [-, -]	-	.62** [-, -]	-	.62** [-, -]
Averill (2016)[f]	851	*M* = 412 days	.33**	- [-, -]	.39**	- [-, -]	-.33**	- [-, -]	.39**	- [-, -]
Wardrop (2020)[g]	483	1 year	-	.69 [.64, .74]	-	.63 [.58, .68]	-	.65 [.60, .70]	-	.69 [.64, .74]
Wardrop (2020)[h]	478	1 year	-	.55 [.50, .60]	-	.55 [.50, .61]	-	.53 [.47, .58]	-	.54 [.50, .60]
General recidivism (excluding supervision violations)										
Hanby (2013)[b]	1,775	2 years	-	.61 [.59, .64]	-	.66 [.63, .68]	-	.61 [.58, .64]	-	.65 [.62, .68]
Hanby (2013)[c]	3,372	2 years	-	.62 [.60, .64]	-	.67 [.65, .69]	-	.62 [.60, .64]	-	.66 [.64, .68]
Chadwick (2014)[d]	391	*M* = 250 days	-	.52 [.42, .69]	-	.53 [.43, .63]	-	.56 [.47, .66]	-	.53 [.43, .63]
Yesberg & Polaschek (2015)[e]	299	6 months	-	.61** [-, -]	-	0.57* [-, -]	-	.60** [-, -]	-	.62** [-, -]
Ferguson (2015)[i]	122	*M* = 809 days	-	.74 [.64, .84]	-	.68 [.56, .79]	-	.65 [.54, .76]	-	.74 [.63, .84]
Ferguson (2015)[j]	122	*M* = 812 days	-	.64 [.53, .76]	-	.63 [.51, .75]	-	.59 [.47, .70]	-	.65 [.54, .76]
Averill (2016)[f]	851	*M* = 412 days	.20**	- [-, -]	.24**	- [-, -]	-.21**	- [-, -]	.24**	- [-, -]
Serin et al. (2016)[k]	563	3 months	-	.67 [.59, .75]	-	.68 [.58, .77]	-	.67 [.59, .75]	-	.70 [.61, .78]
Muirhead (2016)[l]	369	*M* = 973 days	-	.70 [.64, .76]	-	.69 [.63, .74]	-	.68 [.62, .74]	-	.73 [.67, .78]
Perley-Robertson (2018)[m]	112	*M* = 35 months	-	.60 [.49, .70]	-	.62 [.51, .72]	-	.41 [.31, .52]	-	.61 [.50, .71]
Sexual recidivism										
Smeth (2013)[n]	193	*M* = 22 months	$.01_{ns}$	.51 [.26, .75]	$.05_{ns}$	.53 [.30, .75]	$.01_{ns}$	.46 [.21, .71]	$.02_{ns}$	.46 [.22, .70]
Averill (2016)[f]	851	*M* = 412 days	$.02_{ns}$	- [-, -]	.07*	- [-, -]	$-.06_{ns}$	- [-, -]	$.06_{ns}$	- [-, -]
Violent recidivism										
Yesberg & Polaschek (2015)[e]	299	6 months	-	$.58_{ns}$ [-, -]	-	$.55_{ns}$ [-, -]	-	$.61_{ns}$ [-, -]	-	$.60_{ns}$ [-, -]
Averill (2016)[f]	851	*M* = 412 days	.18**	- [-, -]	.26**	- [-, -]	-.17**	- [-, -]	.23**	- [-, -]
Muirhead (2016)[l]	397	*M* = 973 days	-	.61 [.54, .68]	-	.68 [.61, .75]	-	.62 [.55, .68]	-	.66 [.59, .73]
Perley-Robertson (2018)[m]	112	*M* = 42 months	-	.65 [.54, .75]	-	$.60_{ns}$ [.49, .71]	-	.35 [.24, .46]	-	.65 [.55, .76]
Intimate Partner Violence Recidivism										
Perley-Robertson (2018)[m]	112	*M* = 46 months	-	.62 [.50, .73]	-	$.55_{ns}$ [.43, .68]	-	.42 [.29, .55]	-	.60 [.48, .72]
Supervision violations										
Smeth (2013)[n]	193	*M* = 22 months	.33**	.69, [.62, .78]	.27**	.65 [.57, .74]	-.18*	.61 [.52, .69]	.31**	.68 [.60, .76]
Chadwick (2014)[d]	391	*M* = 250 days	-	.61 [.56, .67]	-	.59 [.53, .65]	-	.57 [.51, .63]	-	.61 [.56, .67]
Yesberg & Polaschek (2015)[e]	299	6 months	-	$.53_{ns}$ [-, -]	-	$.53_{ns}$ [-, -]	-	$.58_{ns}$ [-, -]	-	$.55_{ns}$ [-, -]
Ferguson (2015)[i]	101	*M* = 270 days	-	.72 [.62, .81]	-	.70 [.59, .81]	-	.65 [.55, .76]	-	.71 [.60, .81]
Ferguson (2015)[j]	101	*M* = 266 days	-	.68 [.52, .85]	-	.77 [.63, .91]	-	.59 [.46, .71]	-	.74 [.61, .87]
Averill (2016)[f]	851	*M* = 412 days	.31**	- [-, -]	.37**	- [-, -]	-.31**	- [-, -]	.37**	- [-, -]
Serin et al. $(2016)^{l}_{any}$	563	3 months	-	.66 [.60, .72]	-	.68 [.62, .74]	-	.70 [.65, .76]	-	.71 [.65, .77]
Serin et al. $(2016)^{l}_{serious}$	563	3 months	-	.66 [.60, .73]	-	.70 [.63, .77]	-	.70 [.64, .76]	-	.72 [.65, .78]

Perley-Robertson (2018)[m]	112	*M* = 22 months	-	.65 [.54, .76]	-	.65 [.54, .76]	-	.38 [.27, .49]	-	.65 [.54, .76]
Wardrop (2020)[g]	483	1 year	-	.64 [.59, .70]	-	.62 [.56, .68]	-	.63 [.57, .69]	-	.67 [.61, .72]
Wardrop (2020)[h]	478	1 year	-	.56 [.50, .61]	-	.57 [.52, .63]	-	.53 [.47, .56]	-	.56 [.50, .61]

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry; *r* = point-biserial correlations; AUC = area under the curve statistics; 95% CI = 95% confidence intervals; ns = non-significant; "-" = statistic absent. Adapted from Lloyd (2015), Pardoel (2018), and Perley-Robertson (2018). Fourteen additional DRAOR focused studies were omitted due to incompatible or unavailable statistics (Ghartey, 2020; Lloyd, 2015; Lloyd, 2020a; Lloyd, 2020b; Lowenkamp et al., 2016; Pardoel, 2018; Perley-Robertson et al., 2020, Scanlan, 2015; Scanlan, 2020; Serin et al., 2020; Serin & Prell, 2012; Stone, 2017; Tamatea & Wilson, 2009; Yesberg et al., 2015).

[a]Mean follow-up times are reported for studies with variable follow-up periods; all other follow-ups were fixed. [b]General justice involved adult males and females in New Zealand (justice involved Māori individuals only; Hanby 2013).[c] General justice involved adult males and females in New Zealand (overall sample; Hanby 2013). [d]General justice involved adult males and females in Iowa (Chadwick, 2014). [e]High-risk justice involved adult males in New Zealand (Yesberg & Polaschek, 2015). [f]Justice involved adult males who have offended sexually in New Zealand (Averill, 2016). [g]Justice involved adult males and females without a mental disorder in Iowa (Wardrop, 2020). [h]Justice involved adult males and females with a mental disorder in Iowa (Wardrop, 2020). [I]General justice involved adolescent males in New Zealand; AUCs reported here are based on proximal DRAOR scores (Ferguson, 2015). [j]General justice involved adult males in New Zealand; AUCs reported here are based on proximal DRAOR scores (Ferguson, 2015). [k]General justice involved adult males and females in Iowa (Serin et al., 2016). [l]General justice involved adolescent males and females in New Zealand; AUCs reported here are based on proximal DRAOR scores (Muirhead, 2016). [m]Justice involved adult males with a previous, index, and/or recidivistic intimate partner violence offence in Iowa (Perley-Robertson, 2018). [n]Justice involved adult males who have offended sexually in Iowa (Smeth, 2013).

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

Contextualizing the Present Research

In the United States, justice involved individuals are being placed under community supervision at increasing rates (Herberman & Bonczar, 2015). Notably, however, the rate of reincarceration due to supervision violations is also increasing (Grattet et al., 2009).[22] In fact, recent data reported by the Council of State Government (CSG) Justice Centre (2019) indicates that 25% of state prison admissions in the United States are due to technical violations committed while on probation or parole. As explained by the CSG Justice Centre (2019), the high volume of imprisonments due to technical violations has burdened budgets within the criminal justice system because the practice of incarcerating large numbers of justice involved individuals for reasons related to technical violations is considerably more costly compared to cheaper community supervision alternatives. Therefore, from an economic standpoint, it is within the criminal justice system's best interest to investigate more efficient and cost-effective methods for managing justice involved individuals within the community.

As explained by Perley-Robertson (2018), not all tools that are used for the prediction of future risk were specifically designed for this purpose; rather, some tools were designed to assist with case management by informing decision-makers of factors that are relevant to assessing a justice involved person's current risk state. These correctional tools therefore assist with managing recidivism risk by guiding supervision officers in the creation of individualized rehabilitation and risk management plans that are based on the real-time risks and needs of justice involved individuals (Perley-Robertson, 2018). The DRAOR is an example of an emerging case management tool, which was designed to assist CSOs in assessing the current risk states of their clients and designing and optimizing individualized risk management and

[22] Between 1989 and 2009, US rates of reincarceration due to parole violations increased six-fold at the national level and 30-fold in the state of California (Grattet et al., 2009).

community supervision plans. The DRAOR offers the opportunity to capture information about dynamic factors that have been consistently linked to recidivism risk in a more systematic and structured way compared to current case management strategies. By administering the DRAOR at each meaningful point of contact, CSOs can gather relevant information about their client's current risk state and document changes that occur throughout the supervision period.

As discussed above, previous DRAOR validation studies have yielded mixed results. In previous research, the DRAOR has demonstrated varying degrees of predictive accuracy across a wide range of typologies (e.g., general, high-risk, sexual, women, youth, Māori, intimate partner violent, and non-disordered justice involved individuals). Notably, however, the DRAOR does not appear to be predictive of recidivism among justice involved individuals with a diagnosed mental disorder (Wardrop, 2020) and it has not demonstrated an ability to accurately predict intimate partner violent (Perley-Robertson et al., 2020) or sexual recidivism (Averill, 2016; Smeth, 2013). Given these observations, it is important to continue investigating the psychometric properties of the DRAOR.

The primary focus of the present research differs from most previous validation studies on the DRAOR. Like Ghartey (2020), the present study seeks to evaluate whether the DRAOR can function as a risk prediction instrument by employing a weighted approach. Theoretically, introducing and establishing item weight should increase the DRAOR's predictive accuracy by relying on statistical relationships with recidivism rather than judgement. However, Grann and Långström (2007) contend that (1) simple weighting techniques do not yield improved predictive accuracy, and (2) complex weighting procedures merely produce a statistical shrinkage effect (i.e., extreme values "shrink" toward the central value, ultimately reducing sampling and non-sampling errors but potentially adding bias to the model). Reformulating the DRAOR to include

weighted items would result in it more closely following an actuarial approach rather than a structured professional judgement approach. The actuarial approach dictates which information should be considered (i.e., the scale's items and scoring criteria) and the relative importance of such information (i.e., the item weights) in the determination of an individual's probability of reoffending. As previously discussed, Ægisdóttir and colleagues (2006) found that actuarial assessments, on average, perform with 13% greater accuracy compared to assessments that rely on unstructured clinical judgement. It is therefore unsurprising that the benefits of actuarial approaches (e.g., inherent objectivity, greater statistical accuracy) have led some scholars (e.g., Quinsey et al., 1998) to advocate for the complete abandonment of professional judgement in favour of adopting strictly actuarial methods. However, it should also be noted that weighted scales are less generalizable, meaning they may only be applicable to the population for which they were optimized. The previous research by Ghartey (2020) employed a simple weighting technique (i.e., an adapted version of the Burgess method, also known as the simple summation technique; Hakeem, 1948) on the DRAOR. Ghartey (2020) observed that the weighted version of the DRAOR demonstrated slightly higher AUC values compared to the original unweighted DRAOR, although the differences were non-significant. The goal of the present research was to use a more sophisticated weighting approach, with the hope of yielding more promising results.

Study 1: Deriving and Validating DRAOR Item Weights

The purpose of Study 1 was to investigate whether the DRAOR could operate as a weighted risk prediction instrument in a large representative sample of paroled justice involved New Zealanders.

Method

Participants

Participants were drawn from a large dataset (*N* = 3,694) comprised of all justice involved New Zealanders who were paroled over a two-year period (i.e., April 1, 2010 to March 31, 2012). After exclusionary criteria were applied, a final sample of 3,648 cases remained (see below for a more in-depth discussion of the data cleaning process).

In New Zealand, all justice involved individuals who are serving sentences of two or more years are required to report to their CSO for at least six months following their release, with longer periods of supervision for those who are released prior to the expiration of their sentence (Yesberg et al., 2015). Beginning on April 1, 2010, the New Zealand Department of Corrections adopted the DRAOR and mandated that CSOs must apply the DRAOR at each point of "quality contact" (i.e., any meeting with enough time to assess the items) with individuals on community supervision. This dataset reflects the first 24 months of DRAOR implementation in New Zealand, during which CSOs used the DRAOR to assess and reassess all individuals on their caseload on a regular basis (approximately weekly or fortnightly), yielding 97,185 assessments (95,118 once exclusionary criteria were applied). This dataset has been used in multiple prior studies (Hanby, 2013; Lloyd, 2015; Lloyd et al., 2020a; Lloyd et al., 2020b; Stone, 2017).

Apparatus

The current study employed data from baseline and proximal assessments of justice involved individuals being supervised within the community in New Zealand. These assessments were conducted using the Dynamic Risk Assessment for Offender Re-entry (DRAOR; see Appendix A; Serin, 2007, 2015, 2017).

Dynamic Risk Assessment for Offender Re-entry

The DRAOR is a structured case management instrument designed to evaluate stable dynamic risk, acute dynamic risk, and protective factors. The DRAOR follows an interview-based approach that allows CSOs to collect information relevant to the current risk state of justice involved individuals at each point of quality contact. Recall that the DRAOR is composed of 19 items divided into three domains (i.e., Stable, Acute, and Protective). Drawn from the work of Andrews and Bonta (2010) and Hanson and Harris (2000), the six items that compose the Stable domain address attitudes, traits, and behaviours that are relatively enduring, though potentially changeable (Serin et al., 2020). The Acute domain is comprised of seven items that are prone to rapid change, often acting as lifestyle stressors and destabilizers (Serin et al., 2020). Finally, the Protective domain consists of six strength factors which address prosocial perceptions and connectedness (Serin et al., 2020). Importantly, incorporating factors that have been empirically linked to either crime acquisition or desistance into the same model acknowledges that each subset of factors is unique and cannot be defined by the absence or inverse of the other (i.e., risk is not defined by an absence of protective factors, and protection is not defined by an absence of risk factors; Lloyd & Serin, 2012). Put simply, Protective factors should be understood as conceptually independent of risk (Serin et al., 2020).

The DRAOR is theoretically grounded in the Transition Model of Offender Change (R. C. Serin, personal communication, January 20, 2021), which incorporates crime acquisition and desistance factors into a life-course persistence perspective (for an illustration and more fulsome description of the model, see Serin & Lloyd, 2009; Serin et al., 2010). Essentially, the model builds upon Blumstein and Cohen's (1987) age-crime curve by plotting criminal activity over time and strategically placing (1) the processes for crime acquisition and desistance on opposing

ends, and (2) the individual's commitment to change at the plot's apex. The model includes six risk factors for crime acquisition: (1) young age, (2) substance abuse, and the Big Four (i.e., (3) antisocial attitudes, (4) antisocial personality, (5) antisocial history, and (6) antisocial associates; Andrews & Bonta, 2010). The model also includes six desistance correlates: (1) older age, (2) high quality marriage, (3) stable employment, (4) changes in the perception of crime costs versus reward contingencies (i.e., favouring a prosocial lifestyle over remaining criminally active), (5) sobriety from substance use, and (6) prosocial peer associations. Embracing a commitment to change signals the beginning of desistance, but commitment alone is insufficient to sustain the change process. Rather, the change process requires (1) a commitment to change, (2) desistance correlates, and (3) an interplay between internal and external change factors to sustain the intent to change (Serin et al., 2010).

Internal change factors incorporated into the model include (1) agency (i.e., the realistic belief and expectation that change is both possible and likely), (2) attributions (i.e., the use of either internal or external factors to explain events), (3) outcome expectancies (i.e., the anticipated consequences for personal behaviours), (4) identity (i.e., self-perception and how one hopes to be perceived by others), and (5) change beliefs (i.e., the advantages and disadvantages of desisting versus remaining criminally active). External change factors incorporated into the model include (1) correctional intervention (i.e., programming), (2) proactive supervision (i.e., the core correctional practices, such as an appropriate use of authority and a focus on modelling and reinforcement; Andrews and Kiessling, 1980), (3) aftercare (in terms of frequency, quality, and support), (4) positive relationships, and (5) a supportive community. Although it is currently unclear whether an internal switch or an external influence first initiates the transition process, Serin and colleagues (2010) argued that these two forces are likely to operate symbiotically.

Since the DRAOR is composed of dynamic items that have been associated with either reoffending or crime desistance, the DRAOR should theoretically be able to detect changes in risk state over time. As expected, research by Hanby (2013) provided early evidence that changes in DRAOR subscale scores were reflective of changes in risk state. Specifically, Hanby (2013) found that the justice involved individuals who were able to remain crime-free over the two-year assessment period demonstrated significant decreases in their average Stable and Acute subscores and a significant increase in their average Protective subscore. Likewise, using a sample of justice involved adolescents and adults, Ferguson (2015) observed significant decreases in DRAOR Total, Stable, and Acute (sub)scores and significant increases in Protective subscores between first and last assessments.[23] Ferguson (2015) also found that reductions of risk were significantly associated with desistance for both adolescents and adults. Similarly, Muirhead (2016) also found that reductions in risk significantly predicted desistance among justice involved adolescents, observing decreases in DRAOR Total, Stable, and Acute (sub)scores and increases in Protective subscores over time.[24] Finally, Serin and colleagues (2016) observed that 85% of their sample of justice involved individuals on community supervision in Iowa demonstrated changes in their Acute subscores (half increased; half decreased) over their two-month assessment period.[25] Together, these studies provide evidence that suggests that the DRAOR can detect changes in risk state over time.

Given previous research suggesting that the DRAOR is sensitive to changes in risk state over time, some researchers have compared the accuracy of DRAOR assessments at different points of supervision. Lloyd and colleagues (2020a), for example, argued that reassessment

[23] The assessment period was unspecified.
[24] The assessment period was unspecified.
[25] Changes in Stable and Protective subscores were not reported.

improves the DRAOR's predictive validity after observing that (1) reassessments demonstrated incremental prediction over baseline scores, and (2) improved model fit of the most proximal assessment relative to the average of all prior assessments. Additionally, Davies (2019) found consistent evidence that proximal DRAOR assessments are the most accurate predictor of imminent recidivism. Specifically, proximal DRAOR assessments were significantly more accurate predictors of recidivism compared to baseline assessments, and neither aggregation of prior assessments nor various measures of intra-individual change provided a clear improvement of predictive accuracy. Overall, this research further demonstrates that the DRAOR is able to detect changes in risk state over time, with particular emphasis on the merits of proximal scores.

Typically, the DRAOR takes approximately 20 to 30 minutes to score, although timing fluctuations occur based on (1) the complexity of the case, (2) the evaluator's experience, and (3) whether it is an initial or follow-up assessment (Perley-Robertson, 2018). The DRAOR user manual provides question prompts to guide CSOs in conducting interview-style discussions (i.e., there are no specific, standardized questions). Information gleaned from these interviews is used to score each item on a 3-point scale (i.e., 0-2). For items in the Stable and Acute domains, CSOs rate risk factors based on whether they are "not a problem" (i.e., 0), a "slight/possible problem" (i.e., 1), or a "definite problem" (i.e., 2). Protective factors are rated as "not an asset" (i.e., 0), a "slight/possible asset" (i.e., 1), or a "definite asset" (i.e., 2). In the event that information in unreliable or inconsistent, CSOs should aim to resolve uncertainty through additional questioning. An item may be omitted if uncertainty remains, but CSOs are encouraged to consider a score of 1 in such circumstances. Subscores for the Acute domain range from 0-14 and subscores for the Stable and Protective domains both range from 0-12. Total scores are calculated by adding the subscores for the Stable and Acute domains then subtracting the

subscore for the Protective domain, yielding a Total score range from -12 to 26. Theoretically, lower Total scores are expected to be associated with more successful community supervision outcomes (and vice versa).

By conducting a DRAOR assessment at each point of quality contact, CSOs can systematically collect information relevant to their client's current risk state that can be used to create personalized risk management strategies. During follow-up assessments, Stable and Protective item scores only need to be updated when the CSO becomes aware of new, relevant information about the justice involved individual. Score changes on the Stable subscale should reflect recent shifts in engagement with criminal activity (e.g., lowering a score signals the individual's departure from criminal involvement), whereas score changes on the Protective subscale should reflect recent efforts in desistance (e.g., raising a score signals progressive skill-building and engagement in prosocial behaviours). It should take approximately 3-6 months for a justice involved person to demonstrate a one-point change on a Stable or Protective item score (e.g., not an asset to a slight/possible asset), and approximately 6-12 months to demonstrate a two-point change on a Stable or Protective item score (e.g., not an asset to a definite asset). Acute items, on the other hand, must be reconsidered at each new assessment since they are reflective of the current situation and tend to fluctuate rapidly. For these items, changes in scores are based on all relevant behavioural changes since the last assessment, prioritized by recency.

Procedure

On April 1, 2010, the New Zealand Department of Corrections fully implemented the DRAOR for use on all justice involved individuals released on parole. Under their jurisdiction, CSOs are responsible for regularly meeting with all justice involved individuals on their caseload to ensure that the conditions of community supervision are being met. The DRAOR must be

administered at each point of "quality contact" (i.e., any meeting of sufficient length) between the CSOs and their supervisees. During these interviews, CSOs commonly asked questions regarding the justice involved individual's progress in terms of employment, living situation, relationships, and treatment programs (Hanby, 2013).

Alongside a larger, ongoing staff development program, all CSOs were provided with training regarding risk assessments in general, assessment of dynamic risk, and proper use of the DRAOR. This training was delivered through a one-day workshop, which was led by trained instructors who provided opportunities to practice scoring the DRAOR using standardized training materials. Extra supervision was also provided throughout the data collection period to ensure consistency. To address integrity issues, CSOs were able to consult with subject matter experts for the first 12 months of implementation. These subject matter experts received advanced training and monthly supervision from one of the scale developers. CSOs were also provided access to a website containing relevant articles, extra resources, and a DRAOR training centre.

Although the DRAOR was officially fully employed on April 1, 2010, a more sporadic implementation was actually observed as CSOs adjusted to its use in the first few months. Overall, the total number of assessments increased month over month, with 591 assessments in April, 2010 eventually increasing to 5,741 assessments in March, 2012 (Hanby, 2013).[26] Generally, all assessments from April 1, 2010 (i.e., the implementation and study start date) up until either the first reconviction (for recidivists) or the study end date of March 31, 2012 (for non-recidivists) were included in the dataset. Therefore, it is likely that the dataset includes assessments for some justice involved individuals who were already released into the community

[26] The total number of monthly assessments began levelling off at approximately 5,000 per month beginning in March, 2011 (i.e., one year after implementation; Hanby, 2013).

prior to the study start date. Additionally, the dataset also includes assessments for justice involved individuals who were recalled or reconvicted after release, served a short sentence, and then were re-released for a second, or even third, period of parole during the data collection period. These subsequent re-entry trajectories were retained because each period of re-entry may be treated as a theoretically independent unit of analysis as long as the individual's critical information (e.g., their static risk score) was updated upon their return to custody and each baseline assessment was defined by the individual's subsequent release from incarceration (Howard & Dixon, 2013). Follow-up outcome data was collected from April 1, 2010 to July 18, 2012.

Since applying the DRAOR is now part of routine CSO duties in New Zealand, the collection, coding, and entry of all data was completed by staff members of the New Zealand Department of Corrections. Therefore, no recruitment efforts were required. Since the New Zealand Department of Corrections collected this data for client management and program evaluation purposes, informed consent was not obtained and debriefing forms were not administered. This study required identifying information to be able to merge data collected at different time points (i.e., release, community supervision, recidivism, and reconviction). However, to maintain participant anonymity, the data was de-identified and assigned a unique identification number.

Outcome Data

Outcome data took the form of technical violations and any recidivism (defined as any reconviction due to either a parole failure or new conviction) within the data collection period (i.e., April 1, 2010 to July 18, 2012). For criminal recidivism, the date of the offence (rather than the date of detection or return to custody) was recorded, as this method provides the most

accurate timing metric for the outcome behaviour (Lloyd, 2015). It is unclear which date was captured for technical violations, but it is likely that the recorded dates for parole failures reflect a mixture based on (1) the date that the breach occurred, (2) the date that the CSO first became aware of the breach, and (3) the date that the CSO decided to record the breach (Lloyd, 2015). Base rates and average follow-up time in the community prior to an event are presented in Table 2. The follow-up time ranged from 0 days to approximately 2.25 years.

Table 2

*Follow-up Time in Days (M*s and *SD*s) *and Outcome Base Rates*

Outcome	Overall Sample		Construction Subsample		Validation Subsample	
	M (*SD*)	% (*n*)	*M* (*SD*)	% (*n*)	*M* (*SD*)	% (*n*)
Any Recidivism	302.5 (227.0)	42.1 (1535)	308.4 (230.2)	40.2 (734)	296.6 (223.7)	43.9 (801)
Technical Violations	348.7 (235.7)	30.4 (1108)	353.0 (238.5)	29.3 (535)	344.4 (232.8)	31.4 (573)

Note. M = mean. *SD* = standard deviation.

Data Cleaning

This dataset was originally cleaned by Caleb Lloyd for use in his dissertation research (Lloyd, 2015). The following section provides a brief overview of the data cleaning process (for a more fulsome overview, see Lloyd, 2015).

Between April 1, 2010 and March 31, 2012, 3,694 justice involved New Zealanders were released to community supervision. The majority of these individuals (n = 3,498)[27] were assessed with the DRAOR at each point of quality contact with their CSO, yielding a total of 97,188 DRAOR assessments throughout the study period. Three of these assessments contained an incorrect date and were therefore deleted. The data for 11 individuals (totalling 202 assessments) were removed because their release dates were not recorded. Similarly, 190 assessments that

[27] It is unclear why the remaining 196 individuals never underwent a DRAOR assessment. Lloyd (2015) speculates that it is possible that these individuals re-offended and were returned to incarceration before they were able to meet with their CSO for the first time, although he was unable to verify this conjecture.

were recorded early in the implementation period were deleted because they did not have a recorded release date; however, subsequent assessments for these justice involved individuals were retained if they were associated with a later release date. Finally, 66 individuals (totalling 1,675 assessments) were excluded due to issues related to the timing of their assessments. Of these, seven individuals only had assessments that pre-dated their return to the community (it is likely that these assessments were associated with a prior release date that was not recorded in the dataset) and 59 individuals did not receive an assessment within the pre-defined baseline timeframe.[28] Following these exclusions, the final sample was composed of 3,421 justice involved individuals who collectively received a total of 95,118 DRAOR assessments within the study period. Notably, some of these individuals were returned to custody and subsequently re-released during the study period, yielding additional sequences for analysis. Specifically, two sequences were documented for 207 participants and three sequences were documented for 13 participants, yielding a total of 3,654 sequences for analysis. Finally, six sequences were excluded due to missing updated static risk (i.e., ROC*ROI) scores associated with their return to custody. Therefore, the final dataset contained 3,648 sequences for analysis. Only baseline (n = 3,648) and proximal assessments (n = 3,647) were retained for the present research. For a visual depiction of inclusionary criteria, see the flowchart presented in Appendix D.

Analytic Overview

The goal of Study 1 was to derive and validate DRAOR item weights. Therefore, the dataset was randomly divided into two subsamples. The first subsample (i.e., the construction subsample) was used to develop the item weights, and the second subsample (i.e., the validation

[28] To be considered a baseline assessment, a justice involved individual's first assessment must have been completed at any point within their first 28 days within the community.

subsample) was used to test the predictive performance of the item weights. Unless otherwise stated, all analyses were completed using IBM SPSS 27.0.

Given the characteristics of the current dataset, a series of binary logistic regression models were used to develop the DRAOR item weights in the construction subsample. Binary logistic regression is among the most commonly chosen statistical approaches for developing clinical prediction models (van Smeden et al., 2019). Essentially, a binary logistic regression allows researchers to predict which of two categories a person is likely to belong to (e.g., whether or not a justice involved person will violate a supervision condition) when the outcome variable is dichotomous and categorical, and the predictor variables are either categorical or continuous (Field, 2009). Importantly, both of these requirements are satisfied in this dataset. First, the outcome measure (i.e., recidivism) is an example of a dichotomous categorical variable, as it contains two distinct groups that do not have a natural or logical order. Second, the independent predictors (i.e., DRAOR item scores) are all examples of categorical variables, as each item is scored using three pre-defined and logically ordered groups (e.g., not an asset, slight/possible asset, and definite asset) with no quantifiable or necessarily uniform distance between them. Notably, however, the predictor variables were treated as continuous for all analyses to facilitate the current research. Specifically, this study required a single overall weight for each scale item to be able to calculate weighted subscale and Total scores. However, when categorical predictors are included in logistic regression models, the model returns multiple item weights that correspond to each level of the categorical variable rather than a unified weight for the overall variable. Conversely, when continuous predictors are included in logistic regression models, the model produces one weight for the overall variable. For this reason, all DRAOR items were treated as continuous to ensure that the returned item weights could be used to

calculate weighted scores and subscores. Given the primary goal of the current research (i.e., to derive item weights that augment prediction), this was determined to be an appropriate compromise (C. Leth-Steensen, personal communication, May 22, 2021).

Derived item weights were then input into compute statements that were designed to generate weighted DRAOR Total scores and subscores. These weighted scores were used to calculate Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) statistics to determine whether weighting could augment prediction.

Results

Generating Subsamples

The overall sample included 3,648 sequences for analysis, which were randomly divided into two equal subsamples containing 1,824 sequences each. The two subsamples were then randomly assigned to be either the construction or validation subsample.[29] Demographic, offence, and criminogenic risk characteristics are presented in Table 3. Means and standard deviations for DRAOR scores across subsamples are presented in Table 4. All scores were within valid range.

[29] Proximal data was not available for one individual within the validation subsample.

Table 3

Demographic, Offence, and Criminogenic Risk Characteristics Across Subsamples

Indicator	Overall Sample		Construction Subsample		Validation Subsample	
	%	(*n*)	%	(*n*)	%	(*n*)
Age at Release *M* (*SD*)	34.87	(11.55)	35.23	(11.50)	34.52	(11.59)
Ethnicity						
White	37.4	(1363)	39.5	(721)	35.2	(642)
New Zealand Maori	50.0	(1824)	47.4	(864)	52.6	(960)
Pacific Islander	9.9	(360)	10.3	(187)	9.5	(173)
Other	2.8	(101)	2.9	(52)	2.7	(49)
Sex						
Male	92.7	(3383)	92.3	(1683)	93.2	(1700)
Female	7.3	(265)	7.7	(141)	6.8	(124)
Marital Status						
Single	65.9	(2405)	66.3	(1209)	65.6	(1196)
Partnered [a]	16.2	(590)	16.1	(294)	16.2	(296)
Other [b]	17.9	(653)	17.6	(321)	18.2	(332)
Offence Type						
Homicide-Related	4.5	(163)	4.1	(75)	4.8	(88)
Sexual-Related	16.0	(584)	16.8	(307)	15.2	(277)
Robbery	13.9	(508)	13.8	(251)	14.1	(257)
Drug-Related	13.9	(507)	14.4	(263)	13.4	(244)
Assault	8.3	(303)	8.2	(149)	8.4	(154)
Other Violent	12.6	(460)	12.6	(229)	12.7	(231)
Property	26.5	(966)	25.6	(467)	27.4	(499)
Other Non-Violent	4.3	(157)	4.6	(83)	4.1	(74)
ROC*ROI Score *M* (*SD*)	.51	(.24)	.50	(.24)	.52	(.24)

Note. M = mean. *SD* = standard deviation. ROC*ROI = Risk of re-Conviction X Risk of re-Imprisonment model (Bakker et al., 1999). [a]Marital status "partnered" category includes married and common law. [b]Marital status "other" category includes divorced, separated, widowed, unknown, and not specified.

Table 4

DRAOR Descriptives (Ms and SDs) for Items, Subscales, and Total Scores Across Subsamples

DRAOR Variable	Overall Sample		Construction Subsample		Validation Subsample	
	M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Baseline DRAOR Assessment						
Stable	6.39	(2.54)	6.30	(2.56)	6.47	(2.51)
Peer Associations	1.13	(.60)	1.11	(.60)	1.14	(.59)
Attitudes Towards Authority	.87	(.66)	.85	(.66)	.89	(.66)
Impulse Control	1.22	(.57)	1.20	(.58)	1.23	(.57)
Problem Solving	1.17	(.56)	1.16	(.56)	1.19	(.56)
Sense of Entitlement	1.10	(.62)	1.09	(.63)	1.11	(.62)
Attachment with Others	.90	(.57)	.89	(.58)	.91	(.57)
Acute	5.94	(2.42)	5.84	(2.42)	6.03	(2.41)
Substance Abuse	.76	(.67)	.74	(.67)	.77	(.66)
Anger/Hostility	.50	(.62)	.50	(.62)	.50	(.61)
Opportunity/Access to Victims	.90	(.59)	.89	(.59)	.92	(.59)
Negative Mood	.48	(.60)	.47	(.60)	.48	(.59)
Employment	1.13	(.63)	1.12	(.63)	1.14	(.63)
Interpersonal Relationships	1.63	(.64)	1.61	(.66)	1.66	(.62)
Living Situation	.54	(.62)	.51	(.60)	.56	(.63)
Protective	6.03	(2.40)	6.10	(2.36)	5.95	(2.44)
Responsive to Advice	1.02	(.50)	1.04	(.50)	1.01	(.50)
Prosocial Identity	.92	(.53)	.93	(.52)	.92	(.53)
Realistic High Expectations	1.08	(.57)	1.09	(.56)	1.07	(.58)
Cost/Benefits	.99	(.53)	1.01	(.53)	.97	(.54)
Social Support	1.14	(.58)	1.15	(.58)	1.13	(.58)
Social Control	.87	(.50)	.88	(.50)	.85	(.50)
Total	6.30	(6.10)	6.05	(6.04)	6.55	(6.14)
Proximal DRAOR Assessment						
Stable	5.41	(2.99)	5.27	(2.99)	5.54	(2.98)
Peer Associations	1.00	(.68)	.97	(.68)	1.03	(.67)
Attitudes Towards Authority	.67	(.71)	.64	(.70)	.69	(.71)
Impulse Control	1.05	(.67)	1.03	(.68)	1.07	(.67)
Problem Solving	.99	(.66)	.97	(.66)	1.01	(.66)
Sense of Entitlement	.96	(.72)	.95	(.72)	.97	(.72)
Attachment with Others	.74	(.62)	.71	(.61)	.76	(.64)
Acute	4.62	(2.65)	4.52	(2.70)	4.71	(2.60)
Substance Abuse	.49	(.63)	.47	(.63)	.50	(.63)
Anger/Hostility	.29	(.54)	.29	(.54)	.30	(.55)
Opportunity/Access to Victims	.85	(.60)	.83	(.61)	.87	(.55)
Negative Mood	.32	(.55)	.33	(.56)	.31	(.55)
Employment	.96	(.67)	.94	(.67)	.97	(.67)
Interpersonal Relationships	1.31	(.83)	1.27	(.84)	1.35	(.81)
Living Situation	.41	(.61)	.40	(.60)	.42	(.62)
Protective	7.14	(2.80)	7.21	(2.83)	7.08	(2.78)
Responsive to Advice	1.24	(.62)	1.26	(.62)	1.22	(.62)
Prosocial Identity	1.10	(.63)	1.10	(.64)	1.09	(.62)
Realistic High Expectations	1.29	(.63)	1.29	(.63)	1.29	(.63)
Cost/Benefits	1.22	(.64)	1.23	(.64)	1.21	(.63)
Social Support	1.31	(.63)	1.33	(.63)	1.29	(.62)
Social Control	.98	(.57)	1.00	(.58)	.97	(.56)
Total	2.88	(7.35)	2.59	(7.44)	3.17	(7.25)

Note. DRAOR = Dynamic Risk Assessment for Offender Re-entry (Serin, 2007, 2015, 2017). *M* = mean. *SD* = standard deviation.

Data Screening

Logistic regression assumes a linear relationship between the logit transformation of the dependent variable and the continuous predictors (Tabachnick & Fidell, 2013). Following the protocol laid out by Perley-Robertson (2018), this assumption was tested by adding a constant of 13 to DRAOR Total scores to ensure that no values were less than one. Next, the adjusted DRAOR Total and the interaction between the adjusted DRAOR Total and its natural logarithm were added to logistic regression models for each outcome (i.e., technical violations, any recidivism). Since no interaction terms were significant, the linearity of the logit assumption was satisfied for all outcomes. Univariate outliers were examined through standardized residuals. Since no cases had *z*-scores exceeding the absolute critical value of 3.3 (Tabachnick & Fidell, 2013), no values were considered extreme.[30]

Given that the AUC is a non-parametric test, there are no distributional assumptions about the population that need to be examined before conducting these analyses (APA, 2021).

Deriving DRAOR Item Weights

Two different weighting approaches were applied to derive DRAOR item weights for both baseline and proximal assessments using the construction subsample. Although both approaches employed logistic regression, they differed in their restrictiveness. For Approach One, all 19 DRAOR items were entered into the same model simultaneously, as well as age, sex, and ethnicity. As a result, the returned unstandardized beta (*B*) values reflect the slope of the line between the respective item and the outcome of interest (i.e., technical violations or any recidivism), after controlling for age, sex, ethnicity, and the remaining 18 DRAOR items. For Approach Two, 19 distinct models containing age, sex, ethnicity, and one unique DRAOR item

[30] Data screening was only conducted on the construction subsample as it was the only subsample that was analyzed using logistic regression.

were constructed. As a result, Approach Two is less restrictive because the returned unstandardized beta values do not control for the remaining 18 DRAOR items. The unstandardized beta values are presented in Table 5 (technical violations) and Table 6 (any recidivism). A correlation matrix comparing the item weights across weighting approaches is presented in Table 7.

Table 5

Derived DRAOR Item Weights (Technical Violations)

Item	Approach One		Approach Two	
	B	Exp(*B*) [95% CI]	*B*	Exp(*B*) [95% CI]
Baseline Assessment				
Peer Associations	.093	1.098 [.883, 1.364]	.356***	1.428 [1.188, 1.716]
Attitude Towards Authority	.210	1.234 [.985, 1.546]	.368***	1.444 [1.228, 1.699]
Impulse Control	.094	1.098 [.847, 1.423]	.351***	1.421 [1.180, 1.711]
Problem Solving	.102	1.108 [.847, 1.449]	.344***	1.411 [1.167, 1.705]
Sense of Entitlement	-.130	.878 [.694, 1.111]	.240**	1.271 [1.074, 1.504]
Attachment with Others	-.259*	.772 [.617, .964]	.112	1.119 [.934, 1.340]
Substance Abuse	.107	1.113 [.933, 1.328]	.238**	1.269 [1.087, 1.481]
Anger/Hostility	.164	1.178 [.955, 1.454]	.304***	1.356 [1.150, 1.598]
Opportunity/Access to Victims	-.149	.861 [.702, 1.057]	.131	1.140 [.956, 1.360]
Negative Mood	-.137	.872 [.707, 1.074]	.109	1.115 [.939, 1.323]
Employment	.372***	1.451 [1.202, 1.752]	.464***	1.590 [1.341, 1.885]
Interpersonal Relationships	-.037	.963 [.808, 1.149]	.101	1.106 [.937, 1.305]
Living Situation	.180	1.197 [.985, 1.454]	.356***	1.427 [1.204, 1.691]
Responsive to Advice	.251	1.286 [.971, 1.702]	-.352**	.703 [.568, .871]
Prosocial Identity	-.783***	.457 [.341, .612]	-.784***	.457 [.370, .563]
Realistic High Expectations	.428**	1.535 [1.188, 1.983]	-.230*	.794 [.659, .957]
Costs/Benefits	-.311*	.733 [.599, .959]	-.541***	.582 [.476, .712]
Social Support	-.152	.859 [.682, 1.083]	-.370***	.691 [.575, .829]
Social Control	-.011	1.011 [.754, 1.356]	-.511**	.600 [.484, .743]
Constant	.116	1.122 [-, -]	-	- [-, -]
Proximal Assessment				
Peer Associations	.110	1.117 [.906, 1.376]	.530***	1.699 [1.445, 1.998]
Attitude Towards Authority	.081	1.084 [.876, 1.343]	.478***	1.613 [1.391, 1.871]
Impulse Control	-.050	.951 [.761, 1.190]	.447***	1.564 [1.332, 1.836]
Problem Solving	.121	1.128 [.899, 1.417]	.508***	1.661 [1.412, 1.955]
Sense of Entitlement	.031	1.032 [.837, 1.271]	.437***	1.548 [1.335, 1.794]
Attachment with Others	-.166	.847 [.679, 1.056]	.329***	1.389 [1.171, 1.648]
Substance Abuse	.161	1.175 [.971, 1.422]	.491***	1.633 [1.389, 1.920]
Anger/Hostility	.127	1.136 [.892, 1.446]	.562***	1.754 [1.462, 2.104]
Opportunity/Access to Victims	-.175	.839 [.685, 1.028]	.262**	1.300 (1.094, 1.543]
Negative Mood	.103	1.109 [.881, 1.395]	.506***	1.659 [1.388, 1.982]
Employment	.147	1.158 [.964, 1.392]	.447***	1.563 [1.333, 1.832]
Interpersonal Relationships	.186*	1.205 [1.041, 1.394]	.410***	1.506 [1.319, 1.720]
Living Situation	.293**	1.341 [1.102, 1.631]	.607***	1.835 [1.552, 2.169]
Responsive to Advice	-.042	.959 [.758, 1.212]	-.570***	.565 [.477, .670]
Prosocial Identity	-.409**	.664 [.522, .845]	-.717***	.488 [.411, .580]
Realistic High Expectations	.384**	1.469 [1.161, 1.858]	-.364***	.695 [.590, .819]
Costs/Benefits	-.203	.816 [.649, 1.027]	-.592***	.553 [.468, .653]
Social Support	.071	1.074 [.865, 1.334]	-.427***	.653 [.552, .771]
Social Control	-.285*	.752 [.583, .970]	-.704***	.495 [.409, .598]
Constant	-.213	.808 [-, -]	-	- [-, -]

Note. DRAOR = Dynamic Risk Assessment of Offender Re-entry (Serin, 2007, 2015, 2017). All variables were entered simultaneously, controlling for age, sex, ethnicity. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Table 6

Derived DRAOR Item Weights (Any Recidivism)

Item	Approach One		Approach Two	
	B	Exp(*B*) [95% CI]	*B*	Exp(*B*) [95% CI]
Baseline Assessment				
Peer Associations	.067	1.069 [.866, 1.319]	.449***	1.567 [1.314, 1.869]
Attitude Towards Authority	.227*	1.255 [1.010, 1.560]	.517***	1.677 [1.434, 1.961]
Impulse Control	.227	1.255 [.976, 1.614]	.537***	1.710 [1.429, 2.046]
Problem Solving	-.021	.979 [.755, 1.270]	.450***	1.568 [1.306, 1.881]
Sense of Entitlement	-.061	.941 [.749, 1.183]	.412***	1.509 [1.284, 1.775]
Attachment with Others	-.227*	.797 [.641, .991]	.293**	1.340 [1.128, 1.593]
Substance Abuse	.121	1.128 [.951, 1.338]	.354***	1.425 [1.228, 1.654]
Anger/Hostility	.161	1.175 [.958, 1.441]	.463***	1.589 [1.353, 1.867]
Opportunity/Access to Victims	-.112	.894 [.734, 1.090]	.288**	1.333 [1.126, 1.579]
Negative Mood	.050	1.051 [.859, 1.286]	.359***	1.432 [1.214, 1.689]
Employment	.480***	1.616 [1.346, 1.940]	.657***	1.929 [1.634, 2.278]
Interpersonal Relationships	.142	1.153 [.973, 1.367]	.302***	1.353 [1.153, 1.587]
Living Situation	.247*	1.280 [1.057, 1.549]	.528***	1.696 [1.437, 2.003]
Responsive to Advice	.365**	1.440 [1.095, 1.895]	-.416***	.660 [.538, .810]
Prosocial Identity	-.777***	.460 [.345, .612]	-.930***	.395 [.321, .486]
Realistic High Expectations	.173	1.188 [.925, 1.528]	-.480***	.619 [.516, .741]
Costs/Benefits	-.167	.846 [.651, 1.099]	-.592***	.553 [.456, .672]
Social Support	-.161	.851 [.680, 1.066]	-.517***	.596 [.500, .711]
Social Control	-.025	.975 [.733, 1.298]	-.651***	.521 [.423, .642]
Constant	.494	1.639 [-, -]	-	- [-, -]
Proximal Assessment				
Peer Associations	.004	1.004 [.819, 1.230]	.591***	1.806 [1.547, 2.108]
Attitude Towards Authority	.182	1.200 [.973, 1.480]	.618***	1.856 [1.605, 2.146]
Impulse Control	-.058	.944 [.759, 1.174]	.537***	1.711 [1.468, 1.995]
Problem Solving	.064	1.066 [.854, 1.331]	.575***	1.778 [1.521, 2.078]
Sense of Entitlement	.017	1.017 [.830, 1.247]	.519***	1.681 [1.459, 1.937]
Attachment with Others	-.122	.885 [.713, 1.099]	.909***	1.622 [1.375, 1.913]
Substance Abuse	.280**	1.324 [1.094, 1.602]	.693***	2.000 [1.701, 2.351]
Anger/Hostility	.118	1.125 [.880, 1.439]	.723***	2.061 [1.710, 2.485]
Opportunity/Access to Victims	-.150	.860 [.706, 1.049]	.382***	1.465 [1.242, 1.728]
Negative Mood	.077	1.080 [.856, 1.362]	.652***	1.919 [1.605, 2.294]
Employment	.266**	1.305 [1.091, 1.561]	.629***	1.876 [1.607, 2.190]
Interpersonal Relationships	.331***	1.392 [1.212, 1.599]	.582***	1.789 [1.577, 2.030]
Living Situation	.469***	1.598 [1.310, 1.948]	.866***	2.377 [2.001, 2.823]
Responsive to Advice	.100	1.105 [.878, 1.392]	-.613***	.542 [.460, .638]
Prosocial Identity	-.362**	.696 [.551, .880]	-.826***	.438 [.370, .517]
Realistic High Expectations	.188	1.207 [.959, 1.518]	-.585***	.557 [.475, .654]
Costs/Benefits	-.204	.815 [.651, 1.020]	-.705***	.494 [.420, .581]
Social Support	.078	1.081 [.875, 1.335]	-.556***	.573 [.488, .674]
Social Control	-.298*	.743 [.579, .953]	-.824***	.439 [.365, .528]
Constant	.539	1.715 [-, -]	-	- [-, -]

Note. DRAOR = Dynamic Risk Assessment of Offender Re-entry (Serin, 2007, 2015, 2017). All variables were entered into their own model, controlling for age, sex, and ethnicity. * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Table 7

Pearson Correlations Between Item Weights Across Weighting Approaches

Weighting Approach	BT1	BT2	BA1	BA2	PT1	PT2	PA1	PA2
BT1	1.00							
BT2	.358	1.00						
BA1	.835***	.603**	1.00					
BA2	.273	.986***	.551*	1.00				
PT1	.613**	.587**	.677**	.504*	1.00			
PT2	.174	.961***	.439	.977***	.539*	1.00		
PA1	.543*	.600**	.721***	.556*	.902***	.565*	1.00	
PA2	.107	.930***	.391	.963***	.453	.981***	.525*	1.00

Note. BT1 = Baseline prediction of technical violations using Approach 1. BT2 = Baseline prediction of technical violations using Approach 2. BA1 = Baseline prediction of any recidivism using Approach 1. BA2 = Baseline prediction of any recidivism using Approach 2. PT1 = Proximal prediction of technical violations using Approach 1. PT2 = Proximal prediction of technical violations using Approach 2. PA1 = Proximal prediction of any recidivism using Approach 1. PA2 = Proximal prediction of any recidivism using Approach 2. * denotes $p < 0.05$ (2-tailed), ** denotes $p < 0.01$ (2-tailed), *** denotes $p < 0.001$ (2-tailed).

Calculating Weighted DRAOR Scores

For Approach One, weighted DRAOR subscale and Total scores were calculated by inputting the unstandardized beta values into the following SPSS compute statement:

$$\frac{1}{1 + e^{-(B_0 + B_1x_1 + B_2x_2 + \ldots + B_{19}x_{19})}}$$

Given the lack of a unified model with a single intercept in Approach Two,[31] the weighted DRAOR subscale and Total scores were calculated by inputting the unstandardized beta values into the following SPSS compute statement:

$$B_1X_1 + B_2X_2 + \ldots + B_{19}X_{19}$$

[31] A comparison of the two compute statements using the construction subsample and the same unstandardized beta values indicated that, although each statement yielded different weighted scores, the AUC and 95% CI values were unaffected. Additionally, although the two compute statements yielded slightly different weighted scores, these weighted scores were highly intercorrelated. For technical violations, there was a significant near-perfect correlation between Total scores across compute statements for baseline ($r = 0.997$) and proximal assessments ($r = 0.997$). A similar pattern of results was observed for any recidivism, with significant strong correlations for both baseline ($r = 0.973$) and proximal assessments ($r = 0.970$).

Twenty-four DRAOR Total scores were calculated across two outcomes (i.e., technical violations, any recidivism), two time periods (i.e., baseline, proximal), three weighting approaches (i.e., unweighted, Approach One, Approach Two), and two subsamples (i.e., construction, validation). Next, 72 DRAOR subscores were calculated across two outcomes (i.e., technical violations, any recidivism), two time periods (i.e., baseline, proximal), three subscales (i.e., Stable, Acute, Protective), three weighting approaches (i.e., unweighted, Approach One, Approach Two), and two subsamples (i.e., construction, validation). These 96 DRAOR subscale and Total scores were then used to run ROC curve analyses to determine their diagnostic ability. Additionally, four ROC curve analyses were run using the ROC*ROI across two outcomes (i.e., technical violations, any recidivism) and two subsamples (i.e., construction, validation). Collectively, these ROC curve analyses yielded a total of 100 AUC values.

Research Question 1a: Can weighting improve the predictive validity of the DRAOR within the construction subsample?

Research Question 1b: Can weighting improve the predictive validity of the DRAOR within the validation subsample?

Predictive validity results for technical violations are presented in Table 8 and predictive validity results for any recidivism are presented in Table 9. Although no significant differences emerged, the weighted approaches tended to yield slightly higher predictive validity compared to their unweighted counterparts.[32] In terms of general predictive validity, results were promising across all 96 DRAOR trials. Regardless of outcome, (sub)scale, assessment time, weighting approach, and subsample, the DRAOR demonstrated significant predictive validity with small to

[32] Since the unweighted Protective subscale yielded AUC values that predicted desistance from (rather than occurrence of) the outcome of interest, the inverse of these values (i.e., 1-AUC) were used when evaluating all changes in predictive accuracy. Note that the Protective domain automatically shifted outcomes when the weighting procedures incorporated unstandardized beta values associated with recidivism rather than desistance.

large effect sizes and poor to acceptable discrimination.[33, 34, 35] The DRAOR Total generally yielded greater predictive validity than the DRAOR subscales within the same condition, and the DRAOR Acute subscale tended to yield greater predictive validity than the Stable and Protective subscales.[36]

Within the construction subsample, the DRAOR and its subscales significantly predicted technical violations regardless of weighting approach or assessment time (AUC_{Stable} = .589 to .640, AUC_{Acute} = .589 to .663, $AUC_{Protective}$ = .406 to .651, and AUC_{Total} = .615 to .701). At baseline, the unweighted DRAOR consistently yielded the lowest predictive accuracy (AUC_{Stable} = .589, a small effect; AUC_{Acute} = .589, a small effect; $AUC_{Protective}$ = .406, a small effect; AUC_{Total} = .615, a small effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Acute} = .611, a small effect; $AUC_{Protective}$ = .619, a small effect; AUC_{Total} = .663, a moderate effect) and Approach Two (AUC_{Stable} = .597, a small effect).[37] For proximal assessments, the lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Acute} = .659, a moderate effect; $AUC_{Protective}$ = .357, a moderate effect; AUC_{Total} = .666, a moderate effect) and Approach One (AUC_{Stable} = .628, a small effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Acute} = .663, a moderate effect;

---

[33] For forensic research, Rice and Harris (2005) suggest that AUC values above .539 suggest a small effect, values above .639 suggest a moderate effect, and values above .714 suggest a large effect. For AUCs below 0.500, values of .461, .361, and .286 reflect small, moderate, and large effect sizes in the opposite direction.
[34] Hosmer and colleagues (2013) suggest AUC values below .500 suggest no discrimination, values between .500 and .699 suggest poor discrimination, values between .700 and .799 suggest acceptable discrimination, values between .800 and .899 suggest excellent discrimination, and values above .900 suggest outstanding discrimination.
[35] Although the unweighted DRAOR Protective subscale yielded AUC values below .500, this is not an indication of non-discrimination. Rather, this is an indication that the domain predicted desistance from the outcome of interest. Since all AUC values below .500 were between .300 and .500, this suggests poor discrimination.
[36] Non-overlapping confidence intervals suggests a significant difference between two AUCs.
[37] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .594) was used.

AUC_{Total} = .701, a moderate effect) and Approach Two (AUC_{Stable} = .640, a moderate effect; AUC_{Acute} = .663, a moderate effect; $AUC_{Protective}$ = .651, a moderate effect).[38]

Within the construction subsample, the DRAOR and its subscales also significantly predicted any recidivism regardless of weighting approach or assessment time (AUC_{Stable} = .615 to .665, AUC_{Acute} = .637 to .709, $AUC_{Protective}$ = .383 to .672, and AUC_{Total} = .660 to .729). At baseline, the lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Acute} = .637, a small effect; $AUC_{Protective}$ = .383, a small effect; AUC_{Total} = .660, a moderate effect) and Approach One (AUC_{Stable} = .616, a small effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Acute} = .653, a moderate effect; $AUC_{Protective}$ = .623, a small effect; AUC_{Total} = .693, a moderate effect) and Approach Two (AUC_{Stable} = .630, a small effect; $AUC_{Protective}$ = .623, a small effect).[39] For proximal assessments, the lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Acute} = .699, a moderate effect; AUC_{Total} = .700, a moderate effect) and Approach One (AUC_{Stable} = .615, a small effect; $AUC_{Protective}$ = .656, a moderate effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Acute} = .709, a moderate effect; AUC_{Total} = .729, a large effect) and Approach Two (AUC_{Stable} = .665, a moderate effect; $AUC_{Protective}$ = .672, a moderate effect).[40]

Within the validation subsample, the DRAOR and its subscales significantly predicted technical violations regardless of weighting approach or assessment time (AUC_{Stable} = .570 to .650, AUC_{Acute} = .590 to .673, $AUC_{Protective}$ = .417 to .646, and AUC_{Total} = .611 to .686). At baseline, Approach One consistently yielded the lowest predictive accuracy (AUC_{Stable} = .570, a small effect; AUC_{Acute} = .590, a small effect; $AUC_{Protective}$ = .556, a small effect; AUC_{Total} = .611,

[38] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .643) was used.
[39] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .617) was used.
[40] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .666) was used.

a small effect), whereas Approach Two consistently yielded the highest predictive accuracy (AUC_{Stable} = .607, a small effect; AUC_{Acute} = .615, a small effect; $AUC_{Protective}$ = .585, a small effect; AUC_{Total} = .620, a small effect).[41] For proximal assessments, the lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Acute} = .667, a moderate effect; AUC_{Total} = .673, a moderate effect) and Approach One (AUC_{Stable} = .609, a small effect; $AUC_{Protective}$ = .608, a small effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Acute} = .673, a moderate effect; AUC_{Total} = .686, a moderate effect) and Approach Two (AUC_{Stable} = .650, a moderate effect; AUC_{Acute} = .663, a moderate effect; $AUC_{Protective}$ = .646, a moderate effect).[42]

Within the validation subsample, the DRAOR and its subscales also significantly predicted any recidivism regardless of weighting approach or assessment time (AUC_{Stable} = .602 to .678, AUC_{Acute} = .630 to .719, $AUC_{Protective}$ = .393 to .671, and AUC_{Total} = .653 to .729). At baseline, the lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Total} = .653, a moderate effect) and Approach One (AUC_{Stable} = .607, a small effect; AUC_{Acute} = .630, a small effect; $AUC_{Protective}$ = .581, a small effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Total} = .660, a moderate effect) and Approach Two (AUC_{Stable} = .638, a small effect; AUC_{Acute} = .644, a moderate effect; $AUC_{Protective}$ = .608, a small effect).[43] For proximal assessments, the lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Acute} = .711, a moderate effect; AUC_{Total} = .711, a moderate effect) and Approach One (AUC_{Stable} = .602, a small effect; $AUC_{Protective}$ = .636, a moderate effect), whereas the highest predictive accuracy was divided between Approach One (AUC_{Acute} = .719, a

[41] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .583) was used.
[42] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .641) was used.
[43] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .607) was used.

large effect; AUC_{Total} = .729, a large effect) and Approach Two (AUC_{Stable} = .678, a moderate effect; $AUC_{Protective}$ = .671, a moderate effect).[44]

Overall, neither the unweighted DRAOR nor Approach One or Approach Two significantly outperformed the others. Weighting did not yield significantly improved predictive validity compared to the unweighted DRAOR. Although there were no significant differences between weighting approaches, the lowest predictive accuracy was consistently observed in either the unweighted DRAOR or weighting Approach One, whereas the highest predictive accuracy was consistently observed in weighting Approaches One and Two.

[44] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .669) was used.

Table 8

Predictive Validity (AUCs) for Technical Violations Across Two Subsamples of NZ Parolees

Scale	Construction Subsample		Validation Subsample	
	AUC [95% CI]	Cohen's d	AUC [95% CI]	Cohen's d
Baseline Assessment				
ROC*ROI	**.674 [.648, .700]**	.637	**.676 [.651, .701]**	.645
Unweighted DRAOR				
Stable Subscale	**.589 [.560, .617]**	.318	**.601 [.573, .628]**	.361
Acute Subscale	**.589 [.561, .618]**	.318	**.610 [.582, .637]**	.394
Protective Subscale	**.406 [.377, .434]**	.336	**.417 [.389, .445]**	.296
Total	**.615 [.586, .643]**	.413	**.617 [.590, .645]**	.420
Weighted DRAOR (Approach 1)				
Stable Subscale	**.593 [.564, .621]**	.332	**.570 [.542, .598]**	.249
Acute Subscale	**.611 [.583, .640]**	.398	**.590 [.562, .619]**	.321
Protective Subscale	**.619 [.590, .647]**	.428	**.556 [.527, .584]**	.199
Total	**.663 [.636, .691]**	.594	**.611 [.583, .638]**	.398
Weighted DRAOR (Approach 2)				
Stable Subscale	**.597 [.568, .625]**	.347	**.607 [.580, .635]**	.383
Acute Subscale	**.603 [.574, .631]**	.369	**.615 [.587, .642]**	.413
Protective Subscale	**.602 [.574, .631]**	.365	**.585 [.557, .613]**	.303
Total	**.628 [.600, .656]**	.461	**.620 [.593, .647]**	.431
Proximal Assessment				
Unweighted DRAOR				
Stable Subscale	**.634 [.607, .662]**	.484	**.645 [.618, .672]**	.525
Acute Subscale	**.659 [.632, .686]**	.579	**.667 [.640, .693]**	.610
Protective Subscale	**.357 [.331, .384]**	.518	**.359 [.332, .386]**	.510
Total	**.666 [.639, .693]**	.606	**.673 [.647, .700]**	.633
Weighted DRAOR (Approach 1)				
Stable Subscale	**.628 [.600, .656]**	.461	**.609 [.582, .637]**	.391
Acute Subscale	**.663 [.636, .689]**	.594	**.673 [.647, .699]**	.633
Protective Subscale	**.646 [.619, .673]**	.529	**.608 [.580, .635]**	.387
Total	**.701 [.675, .726]**	.745	**.686 [.660, .712]**	.685
Weighted DRAOR (Approach 2)				
Stable Subscale	**.640 [.613, .668]**	.506	**.650 [.623, .677]**	.544
Acute Subscale	**.663 [.637, .690]**	.594	**.672 [.646, .699]**	.629
Protective Subscale	**.651 [.624, .677]**	.548	**.646 [.619, .673]**	.529
Total	**.671 [.645, .698]**	.626	**.677 [.651, .703]**	.649

Note. AUC = Area Under the Curve statistic. NZ = New Zealand. ROC*ROI = Risk of re-Conviction X Risk of re-Imprisonment model (Bakker et al., 1999). DRAOR = Dynamic Risk Assessment of Offender Re-entry (Serin, 2007, 2015, 2017). AUC values below .500 indicate that higher scores (i.e., more Protective factors) were associated with lower rates of the outcome variable of interest. Proximal ROC*ROI scores were not obtained. Cohen's d values were calculated using the conversion tables proposed by Salgado (2017).

Table 9

Predictive Validity (AUCs) for Any Recidivism Across Two Subsamples of NZ Parolees

Scale	Construction Subsample		Validation Subsample	
	AUC [95% CI]	Cohen's d	AUC [95% CI]	Cohen's d
Baseline Assessment				
ROC*ROI	**.729 [.706, .752]**	.862	**.737 [.715, .760]**	.896
Unweighted DRAOR				
Stable Subscale	**.625 [.599, .651]**	.450	**.633 [.608, .659]**	.480
Acute Subscale	**.637 [.611, .663]**	.495	**.642 [.616, .667]**	.514
Protective Subscale	**.383 [.357, .409]**	.420	**.393 [.367, .419]**	.383
Total	**.660 [.635, .685]**	.583	**.653 [.628, .679]**	.556
Weighted DRAOR (Approach 1)				
Stable Subscale	**.616 [.590, .643]**	.417	**.607 [.581, .633]**	.383
Acute Subscale	**.653 [.627, .678]**	.556	**.630 [.604, .656]**	.469
Protective Subscale	**.623 [.597, .649]**	.443	**.581 [.554, .607]**	.289
Total	**.693 [.668, .717]**	.713	**.660 [.635, .685]**	.583
Weighted DRAOR (Approach 2)				
Stable Subscale	**.630 [.604, .656]**	.469	**.638 [.613, .664]**	.499
Acute Subscale	**.644 [.619, .670]**	.598	**.644 [.619, .670]**	.598
Protective Subscale	**.623 [.596, .649]**	.443	**.608 [.582, .634]**	.387
Total	**.666 [.641, .691]**	.606	**.655 [.630, .680]**	.564
Proximal Assessment				
Unweighted DRAOR				
Stable Subscale	**.660 [.635, .685]**	.583	**.674 [.649, .698]**	.637
Acute Subscale	**.699 [.675, .724]**	.737	**.711 [.687, .735]**	.786
Protective Subscale	**.334 [.309, .359]**	.606	**.331 [.306, .356]**	.618
Total	**.700 [.675, .724]**	.741	**.711 [.688, .735]**	.786
Weighted DRAOR (Approach 1)				
Stable Subscale	**.615 [.588, .641]**	.413	**.602 [.575, .628]**	.365
Acute Subscale	**.709 [.686, .733]**	.778	**.719 [.696, .743]**	.820
Protective Subscale	**.656 [.631, .681]**	.567	**.636 [.611, .662]**	.491
Total	**.729 [.706, .752]**	.862	**.729 [.706, .752]**	.862
Weighted DRAOR (Approach 2)				
Stable Subscale	**.665 [.640, .691]**	.602	**.678 [.654, .703]**	.653
Acute Subscale	**.705 [.681, .729]**	.762	**.716 [.693, .740]**	.807
Protective Subscale	**.672 [.647, .697]**	.629	**.671 [.647, .696]**	.626
Total	**.704 [.680, .728]**	.757	**.715 [.692, .739]**	.803

Note. AUC = Area Under the Curve statistic. NZ = New Zealand. ROC*ROI = Risk of re-Conviction X Risk of re-Imprisonment model (Bakker et al., 1999). DRAOR = Dynamic Risk Assessment of Offender Re-entry (Serin, 2007, 2015, 2017). AUC values below .500 indicate that higher scores (i.e., more Protective factors) were associated with lower rates of the outcome variable of interest. Proximal ROC*ROI scores were not obtained. Cohen's d values were calculated using the conversion tables proposed by Salgado (2017).

Research Question 2: Are DRAOR weights subject to significant degradation?

Degradation was assessed by examining 48 trials (e.g., Technical Violation—Baseline—Unweighted—Stable) of predictive validity across subsamples (see Table 10). Although only one

significant difference was observed, interesting trends emerged regarding degradation of predictive validity across subsamples. Both risk subscales performed marginally better in the validation subsample; specifically, the Stable subscale performed slightly better in the validation subsample in eight of twelve trials and better in the construction subsample in four of twelve trials, and the Acute subscale performed slightly better in the validation subsample in seven of twelve trials, equal in one trial, and better in the construction subsample in four of twelve trials. Conversely, the Protective subscale performed marginally better in the construction subsample in all 12 trials. Additionally, the weighted Protective subscale yielded significantly higher predictive accuracy at baseline for technical violations using Approach One in the construction subsample (AUC = .619, a small effect) compared to the validation subsample (AUC = .556, a small effect).[45] No clear pattern emerged for DRAOR Total scores, which performed marginally better in the construction subsample in six of twelve trials, in the validation subsample in five of twelve trials, and equally in one trial.

Although non-significant, approach-level examinations also yielded mixed results. For the unweighted DRAOR, eleven of sixteen trials performed better in the validation subsample and five of sixteen trials performed better in the construction subsample. For Approach One, thirteen of sixteen trials performed better in the construction subsample, two of sixteen trials performed better in the validation subsample, and one trial yielded identical AUC and CI values across subsamples. Finally, for Approach Two, nine of sixteen trials performed better in the validation subsample, six of sixteen values performed better in the construction subsample, and one trial yielded identical AUC and CI values across subsamples.

[45] Importantly, this is consistent with what would be expected by chance. Given a Type I error rate of .05, approximately one out of every twenty analyses would be expected to be significant due to chance. Therefore, this single significant trial (i.e., only 1 out of 48) should not be interpreted as meaningful.

Table 10

Comparison of Trial Predictive Validity Across Subsamples

Trial (Outcome—Assessment Period—Weighting Approach—Scale)	Subsample Yielding Higher AUC
Technical Violation—Baseline—Unweighted—Stable Subscale	Validation
Technical Violation—Baseline—Unweighted—Acute Subscale	Validation
Technical Violation—Baseline—Unweighted—Protective Subscale	Construction
Technical Violation—Baseline—Unweighted—DRAOR Total	Validation
Technical Violation—Baseline—Approach One—Stable Subscale	Construction
Technical Violation—Baseline—Approach One—Acute Subscale	Construction
Technical Violation—Baseline—Approach One—Protective Subscale	**Construction**
Technical Violation—Baseline—Approach One—DRAOR Total	Construction
Technical Violation—Baseline—Approach Two—Stable Subscale	Validation
Technical Violation—Baseline—Approach Two—Acute Subscale	Validation
Technical Violation—Baseline—Approach Two—Protective Subscale	Construction
Technical Violation—Baseline—Approach Two—DRAOR Total	Construction
Technical Violation—Proximal—Unweighted—Stable Subscale	Validation
Technical Violation—Proximal—Unweighted—Acute Subscale	Validation
Technical Violation—Proximal—Unweighted—Protective Subscale	Construction
Technical Violation—Proximal—Unweighted—DRAOR Total	Validation
Technical Violation—Proximal—Approach One—Stable Subscale	Construction
Technical Violation—Proximal—Approach One—Acute Subscale	Validation
Technical Violation—Proximal—Approach One—Protective Subscale	Construction
Technical Violation—Proximal—Approach One—DRAOR Total	Construction
Technical Violation—Proximal—Approach Two—Stable Subscale	Validation
Technical Violation—Proximal—Approach Two—Acute Subscale	Validation
Technical Violation—Proximal—Approach Two—Protective Subscale	Construction
Technical Violation—Proximal—Approach Two—DRAOR Total	Validation
Any Recidivism—Baseline—Unweighted—Stable Subscale	Validation
Any Recidivism—Baseline—Unweighted—Acute Subscale	Validation
Any Recidivism—Baseline—Unweighted—Protective Subscale	Construction
Any Recidivism—Baseline—Unweighted—DRAOR Total	Construction
Any Recidivism—Baseline—Approach One—Stable Subscale	Construction
Any Recidivism—Baseline—Approach One—Acute Subscale	Construction
Any Recidivism—Baseline—Approach One—Protective Subscale	Construction
Any Recidivism—Baseline—Approach One—DRAOR Total	Construction
Any Recidivism—Baseline—Approach Two—Stable Subscale	Validation
Any Recidivism—Baseline—Approach Two—Acute Subscale	Neither (Identical AUCs & CIs)
Any Recidivism—Baseline—Approach Two—Protective Subscale	Construction
Any Recidivism—Baseline—Approach Two—DRAOR Total	Construction
Any Recidivism—Proximal—Unweighted—Stable Subscale	Validation
Any Recidivism—Proximal—Unweighted—Acute Subscale	Validation
Any Recidivism—Proximal—Unweighted—Protective Subscale	Construction
Any Recidivism—Proximal—Unweighted—DRAOR Total	Validation
Any Recidivism—Proximal—Approach One—Stable Subscale	Construction
Any Recidivism—Proximal—Approach One—Acute Subscale	Validation
Any Recidivism—Proximal—Approach One—Protective Subscale	Construction
Any Recidivism—Proximal—Approach One—DRAOR Total	Neither (Identical AUCs & CIs)
Any Recidivism—Proximal—Approach Two—Stable Subscale	Validation
Any Recidivism—Proximal—Approach Two—Acute Subscale	Validation
Any Recidivism—Proximal—Approach Two—Protective Subscale	Construction
Any Recidivism—Proximal—Approach Two—DRAOR Total	Validation

Note. AUC = Area Under the Curve. Subsamples that significantly outperformed their counterpart in the respective trial are bolded.

Research Question 3: Does the DRAOR outperform the ROC*ROI?

Performance was assessed by examining overlapping confidence intervals for AUC values. Although the DRAOR did not significantly outperform the ROC*ROI in any trials, the ROC*ROI significantly outperformed the DRAOR in fifty-nine of ninety-six trials. Overall, proximal DRAOR assessments were more comparable to the ROC*ROI than baseline DRAOR assessments, and mixed results were observed when comparing weighting approaches.

For technical violations, the ROC*ROI demonstrated significantly higher predictive validity in twenty-four of forty-eight trials. At baseline, the ROC*ROI (AUC = .674, a moderate effect) significantly outperformed the DRAOR in nearly all conditions, with only the construction subsample weighted Totals performing at comparable rates ($AUC_{Total\ (Approach\ 1)}$ = .663, a moderate effect; $AUC_{Total\ (Approach\ 2)}$ = .628, a small effect). Results were more promising for proximal assessments, with the DRAOR performing at similar rates to the ROC*ROI (AUC = .676, a moderate effect) in nearly all trials, except for two weighted subscales which were outperformed in the validation subsample ($AUC_{Stable\ (Approach\ 1)}$ = .609, a small effect; $AUC_{Protective\ (Approach\ 2)}$ = .628, a small effect).

For any recidivism, the ROC*ROI demonstrated significantly higher predictive validity in thirty-five of forty-eight trials. At baseline, the ROC*ROI ($AUC_{ROC*ROI}$ = .729, a large effect) significantly outperformed the DRAOR in nearly all conditions once again, with only the construction subsample Approach One weighted Total performing at a comparable rate ($AUC_{Total\ (Approach\ 1)}$ = .693, a moderate effect). For proximal assessments, an interesting pattern emerged in which the ROC*ROI significantly outperformed the Stable and Protective subscores across all trials, whereas Total scores and Acute subscores performed similarly to the ROC*ROI across all trials.

Study 2: Cross-Validating DRAOR Item Weights

The purpose of Study 2 was to investigate the out-of-sample predictive performance of the DRAOR item weights that were derived in Study 1.

Method

Participants

Participants were drawn from the dataset used in the Iowa DRAOR pilot study (Serin et al., 2020). This sample was composed of 510 justice involved men who were serving community supervision orders in the state of Iowa. This dataset has been used in multiple prior studies (Chadwick, 2014; Ghartey, 2020; Perley-Robertson, 2018; Perley-Robertson et al., 2020; Serin et al., 2020).

Apparatus

The current study employed data from baseline risk assessments of justice involved individuals being supervised within the community in Iowa. These assessments were conducted using the Dynamic Risk Assessment for Offender Re-entry (DRAOR; see Appendix A; Serin, 2007, 2015, 2017).

Procedure

CSOs employed by the Iowa Department of Corrections (IDOC) were extended the opportunity to voluntarily participate in the pilot DRAOR implementation study (Serin et al., 2020). There was neither compensation for participation nor penalty for declining. In total, 36 CSOs representing every district in the state of Iowa volunteered to participate (caseload sizes ranged from 9 to 114, with an average of 51; Serin et al., 2020). On August 25, 2010, participating CSOs received a one-day, in-class training session which emphasized the assessment of dynamic risk, the composition of the DRAOR, and scoring methods. This training

session was delivered by the scale's developer, Dr. Ralph Serin. Following the training session, an officer advisory group was established to ensure that trainees would continue to receive the support that they required. Dr. Ralph Serin made himself available to this advisory group so that issues could be addressed as they arose.

Prior to data collection, ethics approval was obtained from Carleton University and the IDOC. After completing training, CSOs began applying the DRAOR during their regularly scheduled supervision sessions. First assessments were completed on all justice involved individuals on the participating CSOs caseloads between March, 2011 and July, 2011.[46] This dataset is composed of initial assessments only. Completed DRAOR assessments were uploaded as field notes into the Iowa Corrections Offender Network, which is a digital environment where the IDOC houses all information pertaining to the justice involved individuals within their jurisdiction. Once assessments were completed, IDOC staff retrieved DRAOR scores, recidivism data, and demographic information.

Outcome Data

Outcome data was coded by Nick Chadwick during the initial evaluation of the pilot dataset for his Master's thesis (Chadwick, 2014). Official records were used to determine whether each justice involved individual incurred any new charges for violating a supervision condition (i.e., technical violations), general recidivism (excluding technical violations), or violent recidivism during the study period.[47] The severity of the first new offence (i.e.,

[46] Since the IDOC implemented the DRAOR across caseloads, the DRAOR was used to assess all individuals on community supervision in the state of Iowa during this period. As a result, the amount of time on supervision prior to the first DRAOR assessment varied between individuals.

[47] Although data for violent recidivism was available, this was not examined in the current study. This decision was made because this outcome measure was not captured in the New Zealand dataset, so appropriate comparisons across samples could not be conducted.

misdemeanour versus felony) was also recorded. For individuals who incurred multiple technical violations, all were compiled to capture the total number of supervision violations in each case, but only the earliest event was retained. For general recidivism, both the presence and type were recorded. In the event of multiple new charges occurring on the same date, only the most severe charge was recorded.

During evaluation of the overall pilot dataset (Serin et al., 2020), the time between initial DRAOR assessments and any subsequent new charge or technical violation was calculated by Nick Chadwick. Follow-up time was determined by calculating the amount of time between each justice involved individual's initial DRAOR assessment and November 16, 2015 (i.e., the pilot study end-date) for non-recidivists, or the date of the first new charge or technical violation for recidivists. Base rates and average follow-up time in the community prior to an event are presented in Table 11. The follow-up time ranged from 0 days to approximately 4.67 years. On average, 7.64 (*SD* = 10.33) violations were recorded throughout each justice involved individual's supervision period.

Table 11

*Follow-up Time in Days (M*s and *SD*s) *and Outcome Base Rates*

Outcome	*M*	(*SD*)	%	(*n*)
Any Recidivism	1254.1	(583.5)	38.04	(194)
Technical Violations	720.1	(738.1)	64.90	(331)

Note. M = mean. *SD* = standard deviation.

Data Cleaning

This pilot dataset was collected during Iowa's implementation of the DRAOR (*N* = 562; Serin et al., 2020). The dataset was cleaned by Nick Chadwick, who examined it for data entry errors and the presence of missing data. There were five cases excluded because they contained

out-of-range DRAOR item values.[48] Aside from one case which was missing ethnicity information, there was no missing data for demographic variables or DRAOR items. The absence of ethnicity data in one case was not concerning in the present study, however, since analyses did not focus on ethnic differences.

Analytic Overview

Study 2 was concerned with using data from a separate jurisdiction to investigate the out-of-sample predictive performance of the DRAOR item weights that were derived in Study 1. Therefore, the item weights derived in Study 1 were input into a compute statement designed to generate weighted scores in the Iowa sample.[49] Next, AUC statistics were calculated to determine whether the previously derived item weights could augment prediction in an independent jurisdiction. Once again, all analyses were completed using IBM SPSS 27.0 unless otherwise indicated.

Results

The sample included 510 cases for analysis. Demographic, offence, and criminogenic risk characteristics are presented in Table 12. Means and standard deviations for DRAOR scores are presented in Table 13. All scores were within valid range.

[48] A cross-comparison of scores was not possible because the original responses were unavailable.
[49] This dataset did not contain proximal assessments. Therefore, only the baseline weights from Study 1 were used.

Table 12

Demographic, Offence, and Criminogenic Risk Characteristics

Indicator	%	(*n*)
Age at Release *M* (*SD*)	32.33	(12.14)
Ethnicity		
White	71.0	(362)
Black	28.8	(147)
Asian/Pacific Islander	0.2	(1)
Sex		
Male	100.0	(510)
Female	0.0	(0)
Marital Status		
Single	59.4	(303)
Partnered [a]	17.6	(90)
Other [b]	22.9	(117)
Offence Type		
Homicide-Related	0.4	(2)
Sexual-Related	23.7	(121)
Robbery	0.8	(4)
Drug-Related	20.6	(105)
Assault	15.9	(81)
Other Violent	4.1	(21)
Property	17.8	(91)
Other Non-Violent	16.7	(85)
IRA Score *M* (*SD*)	12.63	(6.39)

Note. M = mean. *SD* = standard deviation. IRA = Iowa Risk Assessment (Iowa Department of Corrections, 2003). [a]Marital status "partnered" category includes married and common law. [b]Marital status "other" category includes divorced and unknown.

Table 13

DRAOR Descriptives (Ms and SDs) for Items, Subscales, and Total Scores

DRAOR Variable	*M*	(*SD*)
Stable	4.78	(2.66)
Peer Associations	.98	(.55)
Attitudes Towards Authority	.66	(.62)
Impulse Control	.94	(.65)
Problem Solving	.90	(.59)
Sense of Entitlement	.64	(.63)
Attachment with Others	.67	(.56)
Acute	5.40	(2.85)
Substance Abuse	.79	(.76)
Anger/Hostility	.62	(.67)
Opportunity/Access to Victims	.69	(.65)
Negative Mood	.69	(.65)
Employment	.87	(.85)
Interpersonal Relationships	.91	(.62)
Living Situation	.82	(.65)
Protective	6.69	(2.81)
Responsive to Advice	1.17	(.56)
Prosocial Identity	1.09	(.54)
Realistic High Expectations	1.15	(.65)
Cost/Benefits	1.14	(.59)
Social Support	1.16	(.63)
Social Control	.98	(.60)
Total	3.49	(7.43)

Note. DRAOR = Dynamic Risk Assessment for Offender Re-entry (Serin, 2007, 2015, 2017). *M* = mean. *SD* = standard deviation.

Calculating Weighted DRAOR Total Scores

For Approach One, weighted DRAOR subscale and Total scores were calculated using the following SPSS compute statement:

$$\frac{1}{1 + e^{-(B_0 + B_1x_1 + B_2x_2 + \ldots + B_{19}x_{19})}}$$

For Approach Two, the weighted DRAOR subscale and Total scores were calculated using the following SPSS compute statement:

$$B_1X_1 + B_2X_2 + \ldots + B_{19}X_{19}$$

Six DRAOR Total scores were calculated across two outcomes (i.e., technical violations, any recidivism) and three weighting approaches (i.e., unweighted, Approach One, Approach Two).

Next, 18 DRAOR subscale scores were calculated across two outcomes (i.e., technical violations, any recidivism), three subscales (i.e., Stable, Acute, Protective), and three weighting approaches (i.e., unweighted, Approach One, Approach Two). These 24 DRAOR subscale and Total scores were then used to run ROC curve analyses to determine their discriminatory ability. Additionally, two ROC curve analyses were run using the IRA across outcomes. Collectively, these ROC curve analyses yielded a total of 26 AUC values.

Research Question 4: Can derived item weights augment prediction across jurisdictions?

Predictive validity results are presented in Table 14. In terms of general predictive validity, the DRAOR demonstrated significant discrimination in twenty-five of twenty-six trials. For technical violations, the DRAOR demonstrated moderate to large effect sizes and poor to acceptable discrimination. For any recidivism, the DRAOR demonstrated small to large effect sizes and poor to acceptable discrimination. Once again, the DRAOR Total scores generally yielded greater predictive validity than the DRAOR subscores within the same condition, and the Acute subscale tended to yield slightly higher predictive validity than the Stable and Protective subscales. There were no significant differences between DRAOR scores across weighting approaches.

For technical violations, the DRAOR and its subscales demonstrated significant predictive validity regardless of weighting approach (AUC_{Stable} = .655 to .696, AUC_{Acute} = .689 to .718, $AUC_{Protective}$ = .656 to .717, and AUC_{Total} = .723 to .734).[50] The lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Total} = .723, a large effect) and Approach One (AUC_{Stable} = .655, a moderate effect; AUC_{Acute} = .689, a moderate effect; $AUC_{Protective}$ = .656, a moderate effect), whereas Approach Two consistently yielded the highest

[50] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .714) was used.

predictive accuracy (AUC_{Stable} = .696, a moderate effect; AUC_{Acute} = .718, a large effect; $AUC_{Protective}$ = .717, a large effect; AUC_{Total} = .734, a large effect).

For any recidivism, the DRAOR and its subscales demonstrated significant predictive validity in all trials except the weighted Stable subscore under Approach One (AUC_{Stable} = .556 to .558, AUC_{Acute} = .592 to .618, $AUC_{Protective}$ = .572 to .602, and AUC_{Total} = .587 to .609).[51] The lowest predictive accuracy was divided between the unweighted DRAOR (AUC_{Acute} = .592, a small effect; AUC_{Total} = .587, a small effect) and Approach One (AUC_{Stable} = .532, no effect; $AUC_{Protective}$ = .572, a small effect), and the highest predictive accuracy was also divided between the unweighted DRAOR (AUC_{Stable} = .558, a small effect; $AUC_{Protective}$ = .398, a small effect) and Approach One (AUC_{Acute} = .618, a small effect; AUC_{Total} = .609, a small effect).

Overall, neither the unweighted DRAOR nor Approach One or Approach Two significantly outperformed the others. Weighting did not yield significantly improved discrimination compared to the unweighted DRAOR. Weighting yielded mixed results, with no significant differences and no clear pattern emerging. Although Approach Two consistently yielded the greatest predictive accuracy for technical violations, it consistently yielded the middle predictive accuracy for any recidivism. For both outcomes, the poorest predictive accuracy was consistently divided between the unweighted DRAOR and Approach One.

[51] For comparison purposes, the unweighted $AUC_{Protective}$ inverse value (i.e., .602) was used.

Table 14

Baseline Predictive Validity (AUCs) Across Outcomes in a Sample of Iowa Probationers and Parolees

Scale	Technical Violations		Any Recidivism	
	AUC [95% CI]	Cohen's d	AUC [95% CI]	Cohen's d
IRA	**.657 [.608, .707]**	.571	**.588 [.538, .637]**	.314
Unweighted DRAOR				
Stable Subscale	**.688 [.640, .736]**	.693	**.558 [.508, .609]**	.206
Acute Subscale	**.707 [.659, .754]**	.770	**.592 [.543, .641]**	.329
Protective Subscale	**.286 [.238, .333]**	.799	**.398 [.348, .447]**	.365
Total	**.723 [.677, .770]**	.836	**.587 [.537, .636]**	.310
Weighted DRAOR (Approach 1)				
Stable Subscale	**.655 [.604, .706]**	.564	.532 [.481, .583]	.113
Acute Subscale	**.689 [.641, .737]**	.697	**.618 [.569, .668]**	.424
Protective Subscale	**.656 [.603, .709]**	.567	**.572 [.523, .622]**	.256
Total	**.724 [.677, .770]**	.841	**.609 [.559, .658]**	.391
Weighted DRAOR (Approach 2)				
Stable Subscale	**.696 [.648, .743]**	.725	**.556 [.505, .606]**	.199
Acute Subscale	**.718 [.672, .765]**	.815	**.605 [.556, .654]**	.376
Protective Subscale	**.717 [.670, .765]**	.811	**.600 [.551, .650]**	.358
Total	**.734 [.688, .779]**	.883	**.592 [.542, .641]**	.329

Note. AUC = Area Under the Curve statistic. IRA = Iowa Risk Assessment (Iowa Department of Corrections, 2003). DRAOR = Dynamic Risk Assessment of Offender Re-entry (Serin, 2007, 2015, 2017). AUC values below .500 indicate that higher scores (i.e., more Protective factors) were associated with lower rates of the outcome variable of interest. Proximal scores were not obtained. Proximal scores were not obtained. Cohen's d values were calculated using the conversion tables proposed by Salgado (2017).

Research Question 5: Does the DRAOR outperform the IRA?

Performance was assessed by examining overlapping confidence intervals for AUC values. Although no significant differences emerged, the DRAOR demonstrated slightly greater predictive accuracy in the majority of trials. Although the IRA tended to slightly outperform the Stable subscale, the overall DRAOR and the Acute and Protective subscales tended to slightly outperform the IRA across outcomes. For technical violations, the DRAOR outperformed the IRA in the majority of trials. However, the IRA (AUC = .657, a moderate effect) yielded marginally greater predictive accuracy compared to the DRAOR in two conditions ($AUC_{Stable\ (Approach\ 1)}$ = .655, a moderate effect; $AUC_{Protective\ (Approach\ 1)}$ = .656, a moderate effect). For any recidivism, the DRAOR outperformed the IRA in seven of twelve trials. The IRA (AUC = .588,

a small effect) demonstrated slightly higher predictive accuracy compared to the unweighted Stable subscore (AUC = .558, a small effect), the unweighted Total score (AUC = .587, a small effect), the Approach One Stable subscore (AUC = .532, no effect), the Approach One Protective subscore (AUC = .572, a small effect), and the Approach Two Stable subscore (AUC = .556, a small effect).

Discussion

The purpose of the current study was to examine whether applying weighting techniques could augment the DRAOR's discriminatory capacity. This was explored across two studies using data drawn from separate jurisdictions. Although weighting did not significantly improve the DRAOR's discrimination in either sample, the current research nonetheless demonstrated the DRAOR's broad utility as a case management and risk prediction instrument. Importantly, this research served as an independent replication of prior research on the DRAOR's discriminatory capacity using the same two samples (Chadwick, 2014; Ghartey, 2020; Hanby, 2013; Lloyd, 2015; Lloyd et al., 2020a; Lloyd et al., 2020b; Perley-Robertson, 2018; Perley-Robertson et al., 2020; Scanlan, 2015; Serin et al., 2020; Stone, 2017; Yesberg et al., 2015). An overview of the results from the present research is reviewed below.

Overview of Study 1

The first part of the current study involved deriving item weights using logistic regression. Notably, this process involved the decision to treat categorical DRAOR data as continuous due to the lack of a clear alternative. Although unconventional, there are certain circumstances in which categorical data can be treated as continuous. Liu and Agresti (2005), for example, argued that ordinal data may be treated as continuous if the variable in question has an underlying continuous response. This was true in the current study, as each DRAOR item

theoretically exists on a continuum before being reduced to an ordinal score. For example, substance use exists along a complex continuum from experimentation to diagnosable disorder, but the DRAOR conceptualizes substance abuse on a 3-point ordinal scale; realistically, an individual who actively consumes substances may fluctuate anywhere along the substance use continuum, but data regarding their substance use, as captured by the DRAOR, will be ordinal in nature. Second, a key argument presented by those who oppose the practice of treating ordinal data as continuous is that there is no way of determining whether each ordinal category is evenly spaced. However, Pasta (2009) counters this argument by asserting that a similar argument could be made regarding continuous data (e.g., age can be quantified in one year increments, but, depending on the outcome of interest, a one-unit increase in age between a one-year-old and a two-year-old may not necessarily be equivalent to a one-unit increase in age between an eighty-one-year-old and an eighty-two-year-old). Additionally, Pasta (2009) argued that treating categorical variables as continuous is a more powerful approach to analyzing ordinal data, ultimately suggesting that this practice should be considered regularly. Therefore, the decision to treat categorical predictors as continuous requires consideration, but is not necessarily problematic.

The derived item weights were then used to calculate weighted scores in the two New Zealand subsamples. Although two different compute statements were used (due to different restrictions imposed by the two weighting approaches), comparisons across compute statements indicated that scores were highly correlated regardless of which calculation was applied.

Next, the DRAOR's discrimination was assessed by calculating AUC values across weighting approaches, subscales, subsamples, assessment times, and outcomes. Across all 96 trials, the DRAOR demonstrated significant discrimination with small to large effect sizes.

Overall, weighting generally yielded improved discrimination compared to the unweighted DRAOR, though no significant differences emerged. In other words, the DRAOR performed relatively equally, regardless of which weighting approach was or was not applied. Notably, however, the ROC*ROI outperformed the DRAOR in the majority of trials. It is also important to note that the finding that all individual trials were statistically significant could be a consequence of using such a large sample in the current study. Specifically, studies employing large samples (especially those with several hundreds to thousands of participants) have an advantage of considerably higher statistical power, though this carries with it the risk of inflating clinically irrelevant findings to appear statistically significant (Armstrong, 2019; Sullivan & Feinn, 2012). For this reason, conclusions should be drawn from the substantive significance (i.e., effect sizes) of each trial, as effect sizes provide a more meaningful and accurate reflection of the DRAOR's utility compared to simply drawing conclusions from *p* values alone (Sullivan & Feinn, 2012).

Finally, degradation was assessed by comparing discrimination in matched trials across the construction and validation subsamples. Though one significant difference emerged, this was consistent with what would be expected by chance and, therefore, should not be interpreted as meaningful. In other words, the derived item weights performed relatively equally across subsamples.

Overview of Study 2

The goal of Study 2 was to investigate whether the item weights that were derived in Study 1 could sufficiently function in an independent sample. A very similar pattern of results emerged to those in Study 1. Specifically, the DRAOR demonstrated significant discrimination in nearly all trials, though no significant differences emerged across weighting approaches. In

other words, the weights that were derived and tested in the construction and validation (i.e., New Zealand) subsamples performed relatively well in the cross-validation (i.e., Iowa) sample. This is consistent with theoretical expectations, and the same pattern of results has been observed by others who have investigated the out-of-sample predictive performance of actuarial risk assessments (e.g., Johnson et al., 2012). The results of Study 2 also indicated that the DRAOR performed relatively well against the IRA. Specifically, though no significant differences emerged, the DRAOR tended to yield slightly greater discrimination compared to the IRA in the majority of trials.

Notably, the findings of the current study were also consistent with prior research by Ghartey (2020), who found that weighting the DRAOR using the Nuffield (1982) procedure produced slight (i.e., non-significant) increases in AUC values for DRAOR Total scores and subscores using the same sample as the current research. Overall, the results of Study 2 demonstrate the discriminatory capacity of the DRAOR and add to the growing body of literature that supports the continued use of the DRAOR as a risk prediction and case management instrument.

Implications of the Current Research

It was anticipated that introducing item weights would augment the DRAOR's discrimination. Although no significant differences emerged, weighting generally yielded slightly higher AUC values compared to the unweighted DRAOR. That said, the present research adds to a large collection of research that demonstrates the DRAOR's predictive utility regarding general recidivism (Averill, 2016; Chadwick, 2014; Ferguson, 2015; Hanby, 2013; Muirhead, 2016; Perley-Robertson, 2018; Serin et al., 2016; Yesberg & Polaschek, 2015) and supervision violations (Averill, 2016; Chadwick, 2014; Ferguson, 2015; Perley-Robertson, 2018; Serin et al.,

2016; Smeth, 2013; Wardrop, 2020; Yesberg & Polaschek, 2015). Although increases in discrimination were marginal, a finding that was consistent with Ghartey (2020), this line of research is still relatively preliminary and, therefore, worthy of further examination.

Despite observing no significant improvements to discrimination, the present research demonstrated that the practical utility of a scale can be augmented by introducing weighting strategies. Granted that the present study did not apply machine learning techniques, the current findings may still serve as an appropriate backdrop for future research in this area. Specifically, future attempts to improve the DRAOR's discrimination may choose to focus on more sophisticated weighting methods. Though there are several avenues that could be pursued, Grann and Långström (2007) argued that machine learning provides the most sophisticated contemporary method for extricating item weights. It is, therefore, unsurprising that the advent of machine learning has dramatically advanced the field of forensic risk assessment (Berk, 2019), especially given the field's data-driven nature and the ease at which machine learning is able to disentangle and simplify complex and interconnected networks of information (Gibson et al., 2021). As machine learning is increasingly being considered the new frontier of correctional assessments (Gibson et al., 2021), it is important to explore how these emerging techniques could inform research and practice.

Although not a direct focus of the current study, the findings contribute to the literature regarding routine reassessment of dynamic risk. Due to their characteristic propensity to change, dynamic risk factors should be the subject of frequent reassessments (Zamble & Quinsey, 1997). Likewise, Orrick (2012) reasoned that it is essential to monitor intraindividual variations in dynamic risk over time, as doing so can yield significant insight into the course of reintegration and the likelihood of reoffending. Accordingly, it is theoretically expected that criminal

behaviour will be more accurately predicted by risk assessments that immediately precede the transgression. Unsurprisingly, Lloyd and colleagues (2020a) found consistent evidence that re-assessment improves the DRAOR's predictive validity, with proximal DRAOR assessments routinely outperforming assessments that were comparatively distal. Similarly, a general, albeit non-significant, trend was observed in the current research in which proximal assessments tended to yield slightly improved discrimination compared to baseline assessments. This finding corroborates the continued application of the DRAOR at each meaningful point of contact.

Overall, this research provides further evidence that the DRAOR is a viable case management instrument with both risks and strengths contributing to discrimination. Importantly, the overall lack of validated case management tools has been increasingly emphasized by correctional researchers. For example, Bonta and colleagues (2008) argued that many parole officers are ineffective agents of change who poorly adhere to basic rehabilitative principles. This led the researchers to ultimately call for a more systematic and structured training agenda, which, from their perspective, would ultimately increase the quality of support that justice involved individuals receive while on community supervision (Bonta et al., 2008; Bonta et al., 2021). As the DRAOR is a theoretically grounded, desistance-oriented tool that has shown promise as a case management instrument, it holds the potential to begin closing this disparity between theory, training, and practice.

This study also found that the DRAOR demonstrated significant discrimination with small to large effect sizes for technical violations and any recidivism in both the New Zealand and Iowa datasets. It is important, however, to acknowledge the implications of attempting to predict these correctional outcomes. Ostermann and colleagues (2015), for example, argued that our operationalizations of recidivism can impact the effectiveness of reintegration. More

recently, Moore and Eikenberry (2021) found that associations of recidivism outcomes can vary based on how recidivism is operationalized. In other words, the way that recidivism is defined necessarily impacts the results of a study. It is also rare for research in this area to examine the effects of individual CSO characteristics. These are nested data and variation across CSOs could greatly affect base rates. This is particularly relevant regarding the implications of tracking and penalizing technical violations. For example, Bradner and Schiraldi (2020) argued that the New York parole system is plagued by structural racism after finding that Black and Latinx parolees were 5 times and 1.3 times (respectively) more likely to be reincarcerated for technical violations compared to their White counterparts. Similar issues of racial disproportionality have been reported in Iowa as well, where Native American, Hispanic, and African-American parolees are more likely to be reincarcerated due to a technical violation compared to White parolees (Fineran, 2020). This is not to suggest that the DRAOR is inherently biased in its prediction of technical violations, or that we should immediately abandon the notion of technical recidivism; however, it is important to acknowledge and critically engage with such concerns, as any discussion of an instrument that may be used to predict technical violations would be incomplete without acknowledging the limitations inherent in using technical violations as the only outcome variable.

Strengths, Limitations, and Future Directions

As with any research, the present study included strengths and weaknesses that would have undoubtedly impacted the results and conclusions of this research. Several of these strengths and weaknesses were immediately linked to the samples that were used. First, the representativeness of each sample must be considered. The New Zealand sample, for example, was truly representative, as it included data for all individuals under community supervision

across the country during their two-year period of DRAOR implementation. The Iowa dataset, on the other hand, was comparatively less representative. As previously mentioned, the Iowa dataset included all community supervised individuals across the caseloads of 36 CSOs representing every district in the state of Iowa. As a result, this sample is likely to be fairly representative of the Iowan correctional population, though inclusion was solely determined by each CSO's personal choice to voluntarily enrol in the implementation study. Also noteworthy, the Iowa dataset could not be fully representative, as no justice involved women were present. Second, the size of the New Zealand sample should be considered a strength because it provided sufficient power to detect effects that were present, though a by-product of this included the increased risk of inflating *p* values. An additional strength was that the New Zealand dataset allowed for the examination of assessments from two different points in time (i.e., baseline and proximal), which provided an excellent opportunity to showcase the dynamic nature of this instrument. However, it should also be noted that the definition of baseline assessments varied slightly between the New Zealand (i.e., within 30 days pre- or 28 days post-release) and Iowa datasets (i.e., within 60 days of release). Future research should examine the utility of weighting at multiple points throughout supervision (i.e., beyond baseline and proximal) using fully representative samples.

Another strength of the present research was the application of data drawn from two distinct jurisdictions. A common concern within correctional research is the use of data drawn from a single jurisdiction. The generalizability of such research is often limited, as each geographic area is necessarily characterized by their unique pattern of criminal activity, risk assessment and case management practices, and variable recidivism rates. Therefore, the present use of data from two jurisdictions should be interpreted as a strength. However, some differences

between these two datasets challenged the ability to make direct comparisons across jurisdictions in the present research, including the considerably different follow-up times and varying approaches to defining the general recidivism outcome. Therefore, future researchers are encouraged to continue exploring the utility of the DRAOR in a variety of additional jurisdictions.

There were also multiple strengths and limitations regarding DRAOR implementation and data collection in both jurisdictions. First, there was an increased likelihood that justice involved individuals would be missing data if their community supervision term began closer to the date of implementation, as implementation was sporadic in both jurisdictions while CSOs learned to apply the new and unfamiliar instrument (Chadwick, 2014; Lloyd, 2015). Relatedly, an important strength in the current research was that all participating CSOs were trained on the DRAOR, which, theoretically, should have improved the integrity of implementation and, as a result, the calculated predictive accuracy (Flores et al., 2006). However, a limitation is that not all CSOs received the same quality of training. For example, Chadwick (2014) found that scores from formally trained CSOs in Iowa were significantly predictive of general recidivism, though scores from informally trained CSOs were unrelated to general recidivism. This is consistent with research by Hanson and colleagues (2007) that demonstrated that POs who reflected greater conscientiousness yielded more accurate assessments. That said, assessment fidelity appears to have been within acceptable range during implementation in Iowa, although an apparent degradation in assessment fidelity has been observed over the past decade (Wardrop, 2020). Interestingly, this trend directly contrasts findings from Flores and colleagues (2006), who discovered that Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) assessments within the first three years of the scale's implementation in a large Midwestern state

yielded substantially poorer statistical relationships with recidivism compared to LSI-R assessments that were conducted more than three years after the scale's implementation. Next, as explained by both Chadwick (2014) and Lloyd (2015), little is known regarding the potential confounding impact of officer-client relationships because information related to this domain was not collected. Regardless, it is important to acknowledge that the reported results may have been impacted by the supervision style of participating CSOs, their supervision strategies, and the working relationships between CSOs the justice involved individuals on their caseloads (Kennealy et al., 2012). Lastly, the DRAOR's interrater reliability has yet to be assessed, so it is unclear whether the CSOs in either jurisdiction were conducting these assessments with a high degree of consistency. Future researchers are strongly encouraged to begin addressing these limitations. At the very least, the DRAOR should continue to be examined in jurisdictions beyond the period at which it is considered fully implemented, and investigations of interrater reliability would greatly contribute to DRAOR validation efforts. Beyond this, examining the impact of variables related to CSOs (e.g., years of experience) and officer-client relationships would meaningfully inform this literature.

Finally, there were methodological strengths and weaknesses in the current study that must be discussed. Following recommendation from Ghartey (2020), the present research applied logistic regression to see whether a more rigorous analytic approach could provide more optimal item weights. However, this approach likely lost some sophistication given the decision to treat categorical data as continuous under both weighting approaches and given the necessity to use a slightly modified compute statement when deriving weighted subscale and Total scores under the second weighting approach. Future researchers are encouraged to apply additional analytic weighting approaches with varying degrees of complexity to examine alternative

avenues for augmenting prediction. To derive item weights in a highly sophisticated manner, for example, Grann and Långström (2007) recommend building an artificial neural network using multivariate nonlinear regression. Artificial neural networking is an appealing, though markedly complex weighting procedure because of its ability to mimic the problem-solving processes of the human brain (Grann & Långström, 2007). Further, this method is also attractive due to its capacity to remain largely unaffected by the typical issues associated with sparse or problematic datasets (i.e., datasets with many missing values, large measurement error, unknown causality, and/or complex data patterning; Grann & Långström, 2007). Although modern machine learning algorithms are increasingly appearing at the forefront of correctional research (Gibson et al., 2021), simpler decision rules seem to produce heuristics that perform equally well against traditional machine learning algorithms (Jung et al., 2020). Specifically, Jung and colleagues' (2020) select-regress-and-round strategy provides a clear, simplified, and equally effective method for reaching complex decisions in a variety of fields (including bail and parole decision-making scenarios). Under this method, item weights are derived in three simple steps: (1) the full set of each item's features is reduced via forward stepwise regression, (2) the selected features from the first step are run through logistic regression models to obtain fitted coefficients, and (3) the returned coefficients are re-scaled into range and then rounded to the nearest integer (Jung et al., 2020). Promisingly, the authors found their select-regress-and-round strategy to significantly outperform the professional judgement of judges and perform on par with conventional classification algorithms when making judicial pre-trial detention decisions, and this finding replicated across 21 decision-making scenarios outside of the criminal justice realm (Jung et al., 2020).

Conclusion

This study served as the second attempt to derive and validate DRAOR item weights. Consistent with prior research by Ghartey (2020), this study found weighting to marginally improve the DRAOR's discrimination, though no significant differences emerged. Nonetheless, the current research demonstrated that the DRAOR was able to accurately predict technical and general recidivism through its measurement of dynamic risk and protective factors. Overall, the DRAOR appears to be an effective case management instrument that is able to accurately monitor various trajectories of community supervision. Future research could examine training fidelity and competency, as well as increasing or decreasing items to improve discrimination. Keeping the DRAOR as is, future research could examine potential weighting approaches consistent with machine learning approaches, though it is entirely possible that the DRAOR's current composition is its most effective form.

References

Ægisdóttir, S., White, M., Spengler, P., Maugherman, A., Anderson, L., Cook, R., Nichols, C., Lampropoulos, G., Walker, B., Cohen, G., & Rush, J. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341–382. https://doi.org/10.1177/0011000005285875

American Psychological Association. (2021). *APA dictionary of psychology*. Retrieved from https://dictionary.apa.org/

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

Andrews, D. A., & Bonta, J. (2010). The psychology of criminal conduct (5th ed.). LexisNexus/Anderson.

Andrews, D. A., & Kiessling, J. J. (1980). Program structure and effective correctional practices: A summary of the CaVIC research. In R. Ross & P. Gendreau (Eds.), *Effective correctional treatment* (pp. 439-463). Butterworth.

Ara Poutama Aotearoa. (2019). *2018/19 Annual report*. Ara Poutama Aotearoa (New Zealand Department of Corrections). https://www.corrections.govt.nz/__data/assets/pdf_file/0008/38852/Annual_Report_2018_2019_Web_Version_Final.pdf

Armstrong, R. A. (2019). Is there a large sample size problem? *Ophthalmic & Physiological Optics, 39*(3), 129–130. https://doi.org/10.1111/opo.12618

Averill, A. E. (2016). *An investigation of the Dynamic Risk Assessment for Offender Re-entry (DRAOR) with New Zealand sexual offenders* [Unpublished master's thesis]. University of Canterbury.

Babchishin, K. M., & Helmus, L. M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavioral Research, 48*(3), 1021–1031. https://doi.org/10.3758/s13428-015-0627-7

Bakker, L., O'Malley, J., & Riley, D. (1999). *Risk of reconviction: Statistical models predicting four types of re-offending*. Ara Poutama Aotearoa (New Zealand Department of Corrections). https://www.corrections.govt.nz/resources/research_and_statistics/risk-of-reconviction

Berk, R. (2019). *Machine learning risk assessments in criminal justice settings*. Springer International Publishing. https://doi.org/10.1007/978-3-030-02272-3

Blumstein, A. & Cohen, J. (1987). Characterizing criminal careers. *Science (American Association for the Advancement of Science), 237*(4818), 985–991. https://doi.org/10.1126/science.237.4818.985

Bonta, J. (1996). Risk-needs assessment and treatment. In A.T. Hartland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Sage Publications.

Bonta, J., & Andrews, D. (2017). *The psychology of criminal conduct* (6th ed.). Routledge, Taylor and Francis Group.

Bonta, J., Bourgon, G., Rugge, T., Pedneault, C. I., & Lee, S. C. (2021). A system-wide implementation and evaluation of the Strategic Training Initiative in Community

Supervision (STICS). *Journal of Criminal Justice, 74*, 101816–. https://doi.org/10.1016/j.jcrimjus.2021.101816

Bonta, J., Rugge, T., Scott, T.-L., Bourgon, G., & Yessine, A. K. (2008). Exploring the black box of community supervision. journal of offender rehabilitation. *Journal of Offender Rehabilitation, 47*(3), 248–270. https://doi.org/10.1080/10509670802134085

Bradner, K., & Schiraldi, V. (2020). *Racial inequities in New York parole supervision.* Columbia Justice Lab. Retrieved from https://justicelab.columbia.edu/sites/default/files/content/NY%20Parole%20Racial%20Inequities.pdf

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

Chadwick, N. (2020). *Examining trajectories of change on risk and protective factors among white and black men offenders on community supervision in Iowa* [Unpublished doctoral dissertation]. Carleton University.

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Cook, N. (2008). Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry, 54*(1), 17–23. https://doi.org/10.1373/clinchem.2007.096529

CSG Justice Centre. (2019, June 18). *Confined and costly: How supervision violations are filling prisons and burdening budgets.* Council of State Governments. https://csgjusticecenter.org/wp-content/uploads/2020/01/confined-and-costly.pdf

Davies, S. T. (2019). *An investigation of how change in dynamic risk and protective factors affects the prediction of imminent criminal recidivism* [Unpublished doctoral dissertation]. Victoria University of Wellington.

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573

Department of Justice. (2018). Community-based sentencing. Retrieved from https://www.justice.gc.ca/eng/rp-pr/jr/rg-rco/2018/mar07.html

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*(6), 599–610. https://doi.org/10.1037/pas0000475

Ellis, J. L. (2017). Factor analysis and item analysis. In J. L. Ellis (Ed.), *Applying statistics in behavioural research*. Boom. Retrieved from https://www.applyingstatisticsinbehaviouralresearch.com/documenten/factor_analysis_and_item_analysis_version_11_.pdf

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

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage.

Fineran, S. (2020). *Racial disparity report.* Iowa Department of Corrections. Retrieved from https://doc.iowa.gov/sites/default/files/documents/2020/11/final_racial_disparity_report.pdf

Flake, J., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science, 8*(4), 370–378. https://doi.org/10.1177/1948550617693063

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*(5), 523-529. https://doi.org/10.1016/j.jcrimjus.2006.09.007

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15*, 380-387. https://doi.org/10.1037/a0025704

Ghartey, K. E. (2020). *Does weighting improve prediction? Assessing the ability of the Dynamic Risk Assessment for Offender Re-Entry (DRAOR) to function as a risk prediction instrument* [Unpublished honours thesis]. Carleton University.

Gibson, C., Guha, A., & Gash, T. (2021). *Future of criminal justice: Criminal Justice and the technological revolution*. Deloitte Center for Government Insights. Retrieved from https://www2.deloitte.com/global/en/pages/public-sector/articles/criminal-justice-and-the-technological-revolution.html

Grann, L., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior, 34*(1), 22–36. https://doi.org/10.1177/0093854806290250

Grattet, P., Petersilia, J., Lin, J., & Beckman, M. (2009). Parole violations and revocations in California: Analysis and suggestions for action. *Federal Probation, 73*(1), 2–11.

Hakeem, M. (1948). The validity of the Burgess method of parole prediction. *American Journal of Sociology, 53*(5), 376-386. https://doi.org/10.1086/220205

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

Hanson, R. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology = Psychologie Canadienne, 50*(3), 172–182. https://doi.org/10.1037/a0015726

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

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

Herberman, E. J., & Bonczar, T. P. (2015). *Probation and parole in the United States, 2013*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. https://www.bjs.gov/content/pub/pdf/ppus13.pdf

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

Howard, P. D., & Dixon, L. (2013). Identifying Change in the Likelihood of Violent Recidivism: Causal Dynamic Risk Factors in the OASys Violence Predictor. *Law and Human Behavior, 37*(3), 163–174. https://doi.org/10.1037/lhb0000012

Iowa Department of Corrections. (2003). *Iowa classification system – assessment and reassessment of risk: Instructions scoring guide*. Unpublished user-manual.

Jewkes, Y. (2007). *Handbook on prisons*. Willan Publishing.

Johnson, K., Park, K., Bogie, A., Flasch, S., & Cotter, J. (2012). *Developing an actuarial risk assessment to inform decisions made by adult protective services workers*. National Institute of Justice. https://www.ojp.gov/pdffiles1/nij/grants/240410.pdf

Jung, J., Concannon, C., Shroff, R., Goel, S., & Goldstein, D. G. (2020). Simple rules to guide expert classifications. *Journal of the Royal Statistical Society. Series A, Statistics in Society, 183*(3), 771–800. https://doi.org/10.1111/rssa.12576

Kennealy, P. J., Skeem, J. L., Eno Louden, J., & Manchak, S. M. (2012). Firm, fair, and caring officer-offender relationships protect against supervision failure. *Law and Human Behavior, 36*(6), 496-505. https://doi.org/10.1037/h0093935

Klingele, C. (2013). Rethinking the use of community supervision. *The Journal of Criminal Law & Criminology, 103*(4), 1015–1069.

Liu, I., & Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Sociedad de Estadística e Intestigación Operativa Test, 14*(1), 1–73. https://doi.org/10.1007/BF02595397

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694. https://doi.org/10.2466/PR0.3.7.635-694

Lloyd, C. (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* [Unpublished doctoral dissertation]. Carleton University.

Lloyd, C. D., Hanson, R. K., Richards, D. K., & Serin, R. C. (2020a). Reassessment improves prediction of criminal recidivism: A prospective study of 3,421 individuals in New Zealand. *Psychological Assessment*. *32*(6), 568–581. https://doi.org/10.1037/pas0000813

Lloyd, C. D., Perley-Robertson, B., & Serin, R. C. (2020b). Age and Strengths in a Community Corrections Sample. *International Journal of Forensic Mental Health, 19*(3), 215–226. https://doi.org/10.1080/14999013.2019.1684404

Lloyd, C. D., & Serin, R. (2012). Agency and outcome expectancies for crime desistance: measuring offenders' personal beliefs about change. *Psychology, Crime & Law, 18*(6), 543–565. https://doi.org/10.1080/1068316X.2010.511221

Lowenkamp, C. T., Johnson, J. L., Trevino, P., & Serin, R. C. (2016). Enhancing community supervision through the application of dynamic risk assessment. *Federal Probation, 80*(2), 16-20. Retrieved from http://www.uscourts.gov/sites/default/files/80_2_3_0.pdf

McDermott, B., Edens, J., Quanbeck, C., Busse, D., & Scott, C. (2008). Examining the role of static and dynamic risk factors in the prediction of inpatient violence: Variable-and person-focused analyses. *Law and Human Behavior, 32*(4), 325–338. https://doi.org/10.1007/s10979-007-9094-8

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* University of Minnesota Press.

Moore, J., & Eikenberry, J. (2021). Recidivism by conviction offense type and release status among prisoners released in Iowa. *Crime and Delinquency, 67*(3), 344–365. https://doi.org/10.1177/0011128720944075

Muirhead, J. (2016). Risky business: Evaluating the Dynamic Risk Assessment for Offender Re-entry for use with New Zealand youth (Master's thesis). Retrieved from http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/5221/thesis.pdf?sequence=1

Nuffield, J. (1982). *Parole decision-making in Canada.* Solicitor General Canada.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Orrick, E. A. (2012). *Criminal attitudes and reentry: An analysis of recidivism post-release from prison among serious violent offenders.* ProQuest Dissertations Publishing.

Ostermann, M., Salerno, L. M., & Hyatt, J. M. (2015). How different operationalizations of recidivism impact conclusions of effectiveness of parole supervision. *Journal of Research in Crime and Delinquency, 52*(6), 771–796.

Pardoel, K. (2018). *An examination of the influence of race and gender on dynamic risk assessment* [Doctoral dissertation in preparation]. Carleton University.

Parole Board of Canada. (2013). Statutory release and the parole board of Canada – Fact sheet. Retrieved from https://www.canada.ca/en/parole-board/corporate/publications-and-forms/statutory-release-and-the-parole-board-of-canada-fact-sheet.html

Pasta, D. J. (2009). *Learning when to be discrete: Continuous vs. categorical predictors*. ICON Clinical Research. Retrieved from https://support.sas.com/resources/papers/proceedings09/248-2009.pdf

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

Perley-Robertson, B., Serin, R. C., & Chadwick, N. (2020). *Using a general case management tool with partner-violent men on community supervision in Iowa.* Manuscript submitted for publication.

Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., & Thornton, D. (2016). *Static-99R coding rules revised – 2016*. Retrieved from http://www.static99.org/pdfdocs/Coding_manual_2016_v2.pdf

Public Safety Canada. (2007). Assessing the risk of sexual offenders on community supervision: The dynamic supervision project. Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/index-en.aspx#a08

Public Safety Canada. (2019). Corrections and conditional release statistical overview 2018 annual report. Retrieved from: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ccrso-2018/index-en.aspx

Quinsey, V., Harris, G. T., Rice, M., & Cormier, C. (1998). *Violent offenders: appraising and managing risk* (1st ed.). American Psychological Association.

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. https://doi.org/10.1007/s10979-005-6832-7

Salerno, J., Marshall, V. D., & Picken, E. B. (2012). Validity and reliability of the rapid assessment for adolescent preventive services adolescent health risk assessment. Journal of Adolescent Health, 50, 595-599. https://doi.org/10.1016/j.jadohealth.2011.10.015

Salgado, J. F. (2018). Transforming the area under the normal curve (AUC) into Cohen's d, Pearson's r pb, odds-ratio, and natural log odds-ratio: two conversion tables. *The European Journal of Psychology Applied to Legal Context, 10*(1), 35–47. https://doi.org/10.5093/ejpalc2018a5

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

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*(3), 251-270. https://doi.org/10.1177/0093854819896387

Serin, R. C. (2007, 2015, 2017). *The Dynamic Risk Assessment Scale for Offender Re-entry (DRAOR): Guidelines for case planning and risk management using structured assessment of dynamic risk and protective factors.* Unpublished user manual.

Serin, R. C., & Chadwick, N. (2017). *Proposed DRAOR decision rules and case plan strategies.* Unpublished user manual.

Serin, R. C., Chadwick, N., & Prell, L. (2020). *Assessing dynamic risk and protective factors among male probationers and parolees in Iowa: The utility of the Dynamic Risk Assessment for Offender Re-entry.* Manuscript submitted for publication.

Serin, R. C., Gobeil, R., Lloyd, C., Chadwick, N., Wardrop, K., & Hanby, L. (2016). Using dynamic risk to enhance conditional release decisions in prisoners to improve their

outcomes: Dynamic risk and conditional release. *Behavioral Sciences & the Law, 34*(2-3), 321–336. https://doi.org/10.1002/bsl.2213

Serin, R., & Lloyd, C. (2009). Examining the process of offender change: the transition to crime desistance. *Psychology, Crime & Law, 15*(4), 347–364. https://doi.org/10.1080/10683160802261078

Serin, R., Lloyd, C., & Hanby, L. (2010). Enhancing offender re-entry: An integrated model for enhancing offender re-entry. *European Journal of Probation, 2*(2), 53–75. https://doi.org/10.1177/206622031000200205

Serin, R. C. & Prell, L. (2012). *Pathways to crime desistance for probationers.* Paper Presented at American Psychology-Law Society Annual Conference, March, San Juan, Puerto Rico.

Serin, R. C., & Wilson, N. J. (2013, 2017). *Structured Dynamic Assessment Case-management - 21 item (SDAC-21): Guidelines for case management using structured assessment of dynamic risk, responsivity, and protective factors.* Unpublished user manual.

Singh, J., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior, 37*(9), 965–988. https://doi.org/10.1177/0093854810374274

Skelton, A., Riley, D., Wales, D., & Vess, J. (2006). Assessing risk for sexual offenders in New Zealand: Development and validation of a computer-scored risk measure. *Journal of Sexual Aggression, 12*, 277-286. https://doi.org/10.1080/13552600601100326

Smeth, A. (2013). *Evaluating risk assessments among sex offenders: A comparative analysis of static and dynamic factors* [Unpublished master's thesis]. Carleton University.

Smeth, A. (2019). *Validating the Structured Dynamic Assessment Case-Management 21-item (SDAC-21) in a sample of incarcerated offenders* [Unpublished doctoral dissertation]. Carleton University.

Stone, A. G. (2017). Using dynamic risk to predict violent recidivism in "real time": Applying a framework for proximal assessment of risk of general recidivism to predict violent outcomes [Unpublished master's thesis]. University of Texas at El Paso.

Sullivan, G. M., & Feinn, R. (2012). Using effect size–or why the p value is not enough. *Journal of Graduate Medical Education, 4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Tamatea, A., & Wilson, N. (2009). *Dynamic Risk Assessment for Offender Re-entry (DRAOR): A pilot study*. New Zealand Department of Corrections.

Travis, J., & Petersilia, J. (2001). Reentry reconsidered: A new look at an old question. *Crime and Delinquency, 47*(3), 291–313. https://doi.org/10.1177/0011128701047003001

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*(3), 381–390. https://doi.org/10.1037/a0023613

van Smeden, M., Moons, K., de Groot, J., Collins, G., Altman, D., Eijkemans, M., & Reitsma, J. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research, 28*(8), 2455–2474. https://doi.org/10.1177/0962280218784726

Wardrop, K. (2020). *A validation of the Dynamic Risk Assessment for Offender Re-entry (DRAOR) for use with offenders with mental disorder* [Unpublished doctoral dissertation]. Carleton University.

Wilson, C., Desmarais, S., Nicholls, T., Hart, S., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior, 37*(6), 377–388. https://doi.org/10.1037/lhb0000025

Wilson, N. J. (2011). *Release-proposal feasibility assessment-revised*. Unpublished manuscript.

Wong, S., & Gordon, A. E. (1999-2003). *The Violence Risk Scale* [Unpublished manuscript]. University of Saskatchewan.

Yesberg, J. A., & Polaschek, D. 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*(1), 80-99. https://doi.org/10.1080/1068316X.2014.935775

Yesberg, J., Scanlan, J., Hanby, L., Serin, R. C., & Polaschek, D. (2015). Predicting women's recidivism: Validating a dynamic community-based "gender-neutral" tool. *Probation Journal, 62*(1), 33–48. https://doi.org/10.1177/0264550514562851

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

Appendix A

The Dynamic Risk Assessment for Offender Re-Entry (DRAOR)[52]

Stable Risk Factors				
Item	**Score (omit if unknown)**			
Peer Associations	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Attitudes Towards Authority	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Impulse Control	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Problem Solving	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Sense of Entitlement	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Attachment with Others	0 Not a problem	1 Slight/possible problem	2 Definite problem	
			Total Stable Risk Score	

[52] Serin (2007, 2015, 2017). Tables adapted from Perley-Robertson (2018).

Acute Risk Factors				
Item	**Score (omit if unknown)**			
Substance Abuse	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Anger/Hostility	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Opportunity/Access to Victims	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Negative Mood	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Employment	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Interpersonal Relationships	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Living Situation	0 Not a problem	1 Slight/possible problem	2 Definite problem	
			Total Acute Risk Score	

Protective Factors				
Item	**Score (omit if unknown)**			
Responsive to Advice	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Prosocial Identity	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Realistic High Expectations	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Costs/Benefits	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Social Support	0 Not a problem	1 Slight/possible problem	2 Definite problem	
Social Control	0 Not a problem	1 Slight/possible problem	2 Definite problem	
			Total Protective Score	

Appendix B

Proposed DRAOR Cut-Off Values[53]

Risk Level	DRAOR Cut-Off Score
Low	≤ 2
Moderate	3 to 9
Moderate/High	10 to 22
High	≥ 23

[53] First proposed by Serin and Chadwick (2017).

Appendix C

Descriptions of DRAOR Items[54]

Stable Subscale

1. Peer Associations: This item refers to the nature and frequency of associations with criminal individuals. Peers can be partners, family members, friends, or acquaintances with whom the client spends free time. Criminal peers may be those who have committed crime in the past, or would likely accept the client if the client were to commit criminal behavior in the present.
2. Attitudes Towards Authority: This item refers to having a hostile, oppositional, antagonistic, or defiant attitude toward those in authority. Attitude refers to beliefs that authority figures (a) do not deserve to have power over them, (b) do not have a legitimate role to play in keeping order or enforcing rules, (c) do not "play fair" specifically when it comes to the client, or (d) do not deserve respect or consideration from the client. Also, (e) a client may describe avoiding their responsibilities or lying to authority figures as a "game" that everyone plays, suggesting that authority figures do not enforce rules for any higher purpose than to manipulate others and "win" power.
3. Impulse Control: This item refers to either a pattern of (a) the client "doing the first thing that comes into their head" without thinking about the consequences, or (b) the client feeling so overwhelmed by impulses that they tend to give in and behave in ways they say they want to avoid.
4. Problem Solving: This item refers to the client's ability to find solutions to their life problems in a way that takes them away from risk situations and criminal behavior. Good

[54] Obtained from Serin (2007, 2015, 2017).

problem solving involves (a) a desire to find non-criminal solutions to problems, (b) thinking through options to decide on the best prosocial solution, and (c) taking action to make the prosocial solution a reality, which involves following through to find new solutions if setbacks come up in the process. Poor problem solving involves (a) a preference for criminal solutions to problems, (b) remaining inactive rather than engaging in proactive solution- finding, or (c) attempting to fix life problems in a way that causes more serious problems (e.g., puts the individual at risk for new crime, or the solution involves criminal activity).

5. Sense of Entitlement: This item refers to an attitude of self-regard and self-centeredness, at the expense of regard for other's rights. Attitude refers to the clients' personal belief that (a) they deserve to get what they want, no matter the expense to others, (b) others will manipulate, coerce, or deceive them, if they don't manipulate, coerce, or deceive others first, (c) it's only fair that those who fight for their own rights will win out over those who are not as strong, and (d) people who lose out had it coming anyway.
6. Attachment with Others: This item refers to a characteristic, ongoing lack of concern for others, resulting in social disconnection or problematic interpersonal attachments. Poor attachment with others may express as (a) general inattention or indifference to the emotions or needs of others, (b) a callous disregard for the ways others may feel hurt or betrayed by the client's actions, (c) short-term, superficial relationships characterized by opportunistic exploitation, or (d) complete disinterest toward feeling close to or maintaining relationships with others.

Acute Subscale

1. Substance Abuse: This item refers to use of unauthorized substances, including illegal drugs and substances banned by supervision order, and the misuse of other substances, including prescription drugs and alcohol.
2. Anger/Hostility: This item refers to the presence of either (a) "hot" emotions, such as high irritability, exasperation, fury, or rage (for example), or (b) attitudes that support the degradation of others, harm to others, or dehumanization of others. Both (a) anger and (b) hostility result in the client presenting as antagonistic, either by (a) behaving antagonistically, showing signs of a bad temper (i.e., clenched fists, speaking loudly, angry facial expressions), or (b) verbal expressions that others (specific individuals, or groups of individuals) do not deserve fair, ethical, or kind treatment. Unlike sadness (anxiety, depression), anger is focused outward, such that the client reports feeling upset at others (but, both types of emotions can occur simultaneously).
3. Opportunity/Access to Victims: This item refers to the immediate availability of opportunities for crime. This is especially important to consider if the client has history victimizing a preferred victim (either one individual, such as an ex-partner, or a specified group of individuals, such as a sex offender with preference for child victims).
4. Negative Mood: This item refers to the presence of unpleasant emotions, especially agitation, distress, anxiety, stress, or sadness. Unlike the "hot emotions" (anger), the focus of these negative moods is turned inward, such that the client reports feeling unsettled and upset inside of themselves (but both types of emotions can occur simultaneously).

5. Employment: This item primarily assesses whether or not the client is currently employed. Other considerations surrounding employment have additional relevance. Specifically, (a) employability (i.e., does the client have the necessary skills to join the workforce?), (b) engagement (i.e., is the client currently satisfied with existing employment?), and (c) effort (i.e., is the client motivated to gain or maintain employment?).
6. Interpersonal Relationships: This item refers to current problems in close interpersonal relationships in the client's life (which can include antagonism, victimization, breakdown, disconnection, social pressure to engage in criminal activity, etc.). The primary consideration should focus upon romantic partnerships, but close other relationships (with family, housemates, business partners, etc.) deserve consideration in the absence of a romantic partnership, or when there is severe distress or the breakdown of a prosocial relationship that is important to the client.
7. Living Situation: This item primarily assesses whether or not the client is currently living in stable, long-term housing. Stable housing can be considered on a continuum from lack of any suitable housing, or homelessness (a definite problem), temporary or possibly problematic housing situations, such as residence at a halfway house, or "couch surfing" (a possible problem), to safe, suitable, long-term housing (not a problem). However, residing at a halfway house may be rated as definite problem if the client should currently be putting in effort to find other suitable housing, but is not doing so. Alternately, living in a supported residence may be rated as not a problem if the current situation may serve as a stable long- term solution, or the client has already made plans to enter safe, stable housing after leaving the residence.

Protective Subscale

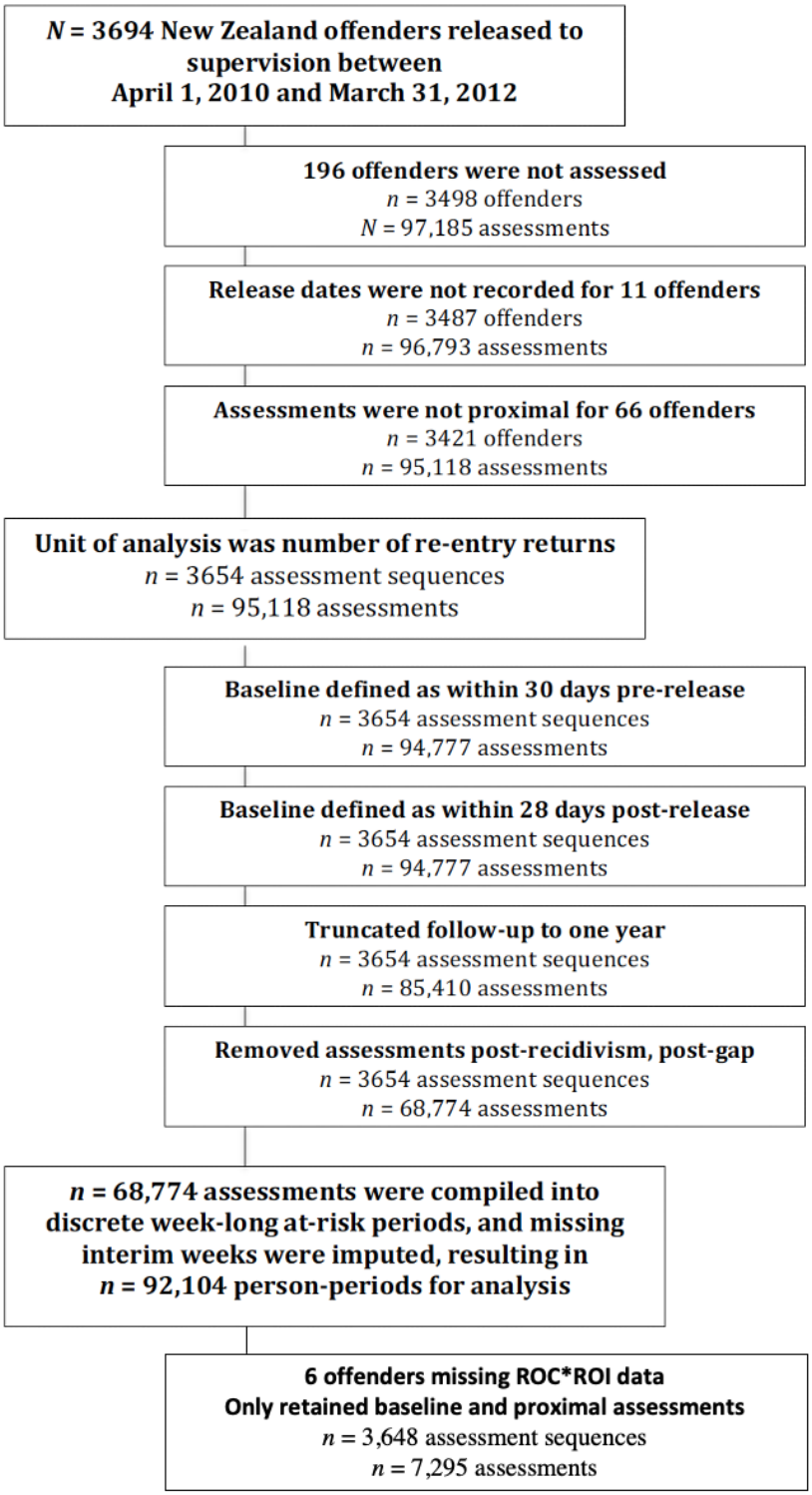
1. Responsive to Advice: This item refers to the client expressing openness to receive and take guidance for making lifestyle changes that will lead toward a long-term, crime-free lifestyle. This is the "Do I care, do I listen, and do I act?" component of the beliefs that support a process of desisting from crime.
2. Prosocial Identity: This item refers to the client's internal self-image. This is the "Who am I?" component of the beliefs that support a process of desisting from crime. The purpose of this item is to assess whether the client can imagine and articulate a "future self" that feels comfortable, fulfilled, and satisfied in a fully non-criminal lifestyle.
3. (Realistic) High Expectations: This item refers to clients' attitude toward change. Attitude refers to two specific beliefs about change that are particularly relevant. First, do clients have a high sense of hope that they will be able to make the prosocial changes they desire to make in their lives? Second, do they believe that change will take time and require personal effort, such that they need to work at creating personal change and re-gaining the trust of others? Combining these two beliefs, do they have a sense of hope that they have the skills it will take to work to achieve the non-criminal life they want?
4. Costs/Benefits (Supportive of Staying Crime-free): This item refers to clients' attitude toward the personal value of crime, and how this attitude compares or contrasts with their attitude toward the value of extending effort to stay crime-free. For this item, attitude refers to beliefs that bad outcomes will happen if choices are made to commit crime in the future, but good outcomes will happen if effort is extended toward staying crime-free. For example, the client who is characterized by prosocial costs/benefit beliefs can articulate that (a) they will lose something important to them if they ever commit another

crime, because the consequences of crime have high drawbacks to them, and (b) they will gain something important to them if they put effort into staying crime-free, even if it is difficult or unfamiliar to them at first.

5. Social Support: This item refers to whether clients have any meaningful relationships with non-criminal individuals, especially individuals who assist the client by offering relational (and/or material) support to the client. This item refers to the nature and frequency of associations with prosocial individuals. These individuals can be partners, family members, friends, or acquaintances with whom the client spends free time. A key feature that identifies a prosocial individual is that they would be unlikely to view the client favorably if the client were to commit new criminal behavior.
6. Social Control: This item refers to the effects that existing prosocial relationships are having on clients who have supportive, non-criminal individuals in their lives. This item extends beyond the simple existence of prosocial others in the client's life, and refers to the influence these individuals have on the client.

Appendix D

Flowchart of Inclusionary Criteria[55]



[55] Adapted from Lloyd (2015).