An Examination of the Influence of Gender and Race on Dynamic Risk Assessment

by

Kaitlin Pardoel

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

Doctor of Philosophy
in
Psychology

Carleton University
Ottawa, Ontario

© 2020
Kate Pardoel

Abstract

Women represent one of the fastest-growing correctional populations and individuals belonging to historically marginalized groups are dramatically overrepresented in prison. Yet, these correctional subpopulations remain poorly understood. Hence, the need for continued research on justice involved (JI) women and individuals belonging to marginalized groups is increasingly being recognized by correctional agencies and researchers, and a body of literature examining the intersection of gender, race, and correctional outcomes has begun to develop. The goal of this dissertation was to contribute to this emerging body of knowledge by exploring the impact of individuals' gender and race on dynamic risk assessment using the Dynamic Risk Assessment for Offender Re-Entry (DRAOR). Study 1 used a sample of 3,091 racially diverse (78.2% White, 18.1% Black, and 3.7% Hispanic) JI women supervised in the community in Iowa to explore the psychometric properties and factor structure of the DRAOR. The DRAOR's original structure provided a poor fit for the data, and alternative models were identified using factor analytic techniques and exploratory structural equation modeling (ESEM). Next, measurement invariance testing was used to determine whether DRAOR scores could be meaningfully compared across race and across assessment occasions. Contrary to expectations, the DRAOR did not demonstrate adequate measurement invariance and subsequent analyses were conducted to identify any item(s) that might explain these findings. Results of these analyses revealed several potential issues and directions for future research are discussed. Using the same sample, Cox regression survival analysis was used to explore whether recidivism rates differed across race in Study 2. Three recidivism outcomes, technical violations, new offences, and any return

were examined. Black women had the highest rates of technical violations and returns, but White women were most likely to commit a new offence. Likelihood of failure was lowest for Hispanic women across all three types of recidivism. Findings from Study 2 also indicated that survival time was significantly associated with women's level of risk, though not always in the expected direction. Further research examining how risk informs the supervision practices employed with JI women is necessary. Study 3 used a subsample of 2,763 women and matched sample of 2,763 men to explore the predictive ability of the DRAOR across gender and race. Prediction was assessed by means of discrimination, calibration, and regression analyses. Despite considerable variability across analyses and type of recidivism examined, findings collectively indicated that DRAOR scores predicted most accurately for White men. Discrimination analyses found that DRAOR scores accurately differentiated between recidivists and non-recidivists for White and Hispanic men and women for the technical violations and any return outcome but were not able to do so for Black justice involved persons (JIPs). Calibration analyses revealed that the DRAOR tended to under-predict technical violations and returns for all JIPs and over-predict new offences, regardless of race and gender. However, the extent of over- and under-prediction was more pronounced for women compared to men. Item-level regression analyses indicated that few DRAOR items reliably predicted recidivism, and analyses regarding potential promotive and protective effects of purportedly women-salient factors assessed by the DRAOR produced inconclusive results. Collectively, findings suggest that parole officers should be cautious in using the DRAOR with JI women and individuals from marginalized groups and underscores the need for additional research examining the influence of gender and race on dynamic risk assessment.

Acknowledgements

I would like to acknowledge the following people for their contributions:

Dr. Ralph Serin, my graduate supervisor. I am grateful for your mentorship over the course of the last 8 years, and for all of the support, advice, understanding, and opportunities you have provided. Your unwavering enthusiasm for the projects at hand and your sense of humour have been greatly appreciated. Thank you.

My committee members, Dr. Shelley Brown and Dr. Julie Blais. Your constructive feedback during the developmental stages of this project helped to shape the final product and pushed me to step out of my academic comfort zone. Thank you.

Dr. Tonia Nicholls, my external examiner, and Dr. Diana Majury, my internal examiner. I sincerely appreciate the time and effort you put into carefully reviewing a very lengthy document. Your insightful comments helped improve the quality of this project. Thank you.

The Iowa Department of Corrections. This project would not have been possible without the collaboration of the IDOC. To all those involved in collecting the data and preparing the datasets, thank you.

Dr. Cecilia Jorgenson, my go-to stats guru. Your assistance throughout the analysis phase of this project was invaluable. From being constantly available to act as a sounding board to letting me take advantage of the expensive statistical software on your computer to run analyses, you helped make this undertaking less dauting. Thank you.

Finally, my partner, family, and friends. You know who you are. I am sincerely grateful for all of the support and encouragement you have offered throughout my university career. Thank you.

Table of Contents

List of Tables

List of Figures

List of Appendices

An examination of the influence of race and gender on dynamic risk assessment

Project Overview

The overarching goal of this dissertation is to promote the integration of the heretofore largely parallel disciplines of dynamic risk assessment, case management, and desistance research. More precisely, this project aims to address risk assessment, risk management, and desistance attending specifically to the influence of gender and race, two areas that have historically received limited attention within correctional research.

At present, there is an abundance of risk assessment instruments designed and validated for use with specific correctional populations. While there are merits to the existence of diverse assessment instruments tailored for use with specific subpopulations of justice involved persons (JIPs), the number and quality of these scales is now such that attempting to create new scales with fractionally improved predictive capabilities yields diminishing returns. Accordingly, this dissertation supports a shift away from this competitive endeavor, toward a concerted focus on the refinement of promising existing instruments and for an empirically grounded exploration of how these instruments can be employed to maximize their efficacy. Once individual instruments' capabilities have been delineated, focus can then shift to considering *when*, with *whom*, and *how* (e.g., in conjunction with which other tools) different assessment instruments might be used to improve predictive accuracy throughout the sentencing process, thereby contributing to the advancement of offender management practices more broadly.

The Dynamic Risk Assessment for Offender Re-Entry (DRAOR; Serin, 2007) is one of the "promising existing instruments" referred to above and will be used to frame the discussion of risk assessment, dynamic risk, and recidivism prediction throughout this

dissertation. Briefly, the DRAOR was developed for use by probation and parole officers, and takes into account Stable, Acute, and Protective factors. It was designed to contribute to effective supervision practices by providing probation officers with a means to assess, and routinely reassess, clients' criminally relevant circumstances thereby allowing for amendment or additions to the case management plan currently in place.

This dissertation centers on an examination of risk and protective factors in a large, racially diverse sample of justice involved (JI) women. Given the scope of this project, the document is broken down into a series of chapters, each with its own specific focus. Chapter 1 provides an overview of relevant theory and literature describing dynamic risk and protective factors, risk assessment, and recidivism prediction in the correctional context to contextualize the later discussion of risk and prediction as it pertains to JI women and historically marginalized groups. Chapter 2 considers the influence of race and gender more specifically and explores the concept of intersectionality. Contemporary issues such as racial disproportionality, possible bias within the correctional system, and the debate surrounding gender-neutrality, gender-salience, and gender-specificity are presented. Chapter 3 reviews relevant research on the DRAOR, including prior validation and implementation research conducted internationally. Chapter 4 details aspects of the methodology that are consistent across the three main studies comprising this dissertation to minimize repetition. These studies are presented in Chapters 5, 6, and 7. As the relevant context and rationale underlying these studies was presented in Chapters 1-3, Chapters 5-7 prioritize discussion of the chosen analytic approach, results, and a discussion of findings. Moreover, because the analytic approach adopted in dissertation is relatively novel (i.e., has not been applied to

this specific population of JIPs and/or the approach differs from traditional approaches), considerable effort was taken to ensure that all methodological decisions and stages of the analytic process are clearly documented. As noted by Earp and Trafimow (2015), reliable, replicable, findings are paramount in research; accordingly, pains were taken to provide detailed information about why and how analyses were carried out. The final chapter in this dissertation, Chapter 8, summarizes and integrates the results of Chapters 5, 6, and 7, and situates the collective findings within the context of the broader correctional literature.

As just alluded to, the research questions examined in this dissertation have been broken down into three distinct studies. The first study consists of a comprehensive examination of the psychometric properties of the DRAOR. To do this, Study 1 explores the DRAOR's underlying factor structure in a sample of racially diverse JI women and evaluates the stability of this factor structure across groups of these women with different racial/ethnic backgrounds as well as over time via measurement invariance testing. The second study establishes base rates for different types of recidivism and plots survival curves for these women. The focus of the final study is prediction. Study 3 explores the impact of gender and race on various aspects of the DRAOR's predictive abilities using matched samples of JI men and women. Overall predictive ability is assessed using discrimination and calibration analyses and logistic regression is used to further explore the impact of intersectionality at an item-level. An examination of strengths and how possible promotive and protective factors function for JI women is also included.

Chapter 1: The Role of Dynamic Risk and Protective Factors

Contextualizing Risk

Continued research investigating the efficacy of contemporary risk assessment instruments is paramount. The growth of prison populations represents a decades-old problem plaguing correctional systems worldwide (Mauer, 1994), and remains a salient issue in North America (Kaeble & Glaze, 2016; Public Safety Canada, 2015), and in the United States in particular (Carson, 2015; Flores, Lowenkamp & Bechtel, 2016; Wagner & Walsh, 2016). Notwithstanding the small improvements seen in the past few years, continued advancements in the realm of risk assessment, an integral component within the larger offender management and reintegration processes, stands to yield substantive benefits. For example, extending the literature in this area could inform correctional policy, guide improvements to the management of resources, and contribute to the promotion of public safety.

While the integral role of risk assessment tools in offender management, reintegration, and continued risk mitigation has been widely acknowledged, there is still considerable disagreement concerning which assessment instruments are most appropriate for use with which subpopulations of JIPs, and about which systemic approaches to assessment (e.g., the desirability of frequent reassessment) offer the greatest advantages. For more information about selecting the ideal assessment tool, readers are encouraged to review Yang, Wong, and Coid's (2010) meta-analytic findings on the subject. The field would benefit from increased clarity regarding the predictive capabilities of existing instruments and from a better understanding of how and when these instruments should be used in order to achieve maximal predictive accuracy.

Assessing Risk

Viewed in its entirety, the criminological literature unequivocally supports the predictive utility of static risk factors (i.e., those factors such as age, gender, and criminal history that cannot be deliberately changed or altered through interventions; Gendreau, Little, & Goggins, 1996). Conversely, dynamic risk factors, which can, in the broadest sense, be conceptualized as potentially changeable factors (e.g., substance use, or antisocial peers), received much less attention historically and were often a subject of contention (Gendreau et al., 1996). That said, recent decades have witnessed a considerable increase in the interest paid to dynamic factors. More recent research (e.g., Brown, St. Amand, & Zamble, 2009; Douglas & Skeem, 2005; Lowenkamp, Johnson, Trevino & Serin, 2016; Serin, Chadwick, & Lloyd, 2015; Serin, Gobeil, Lloyd, Chadwick, Wardrop, & Hanby, 2016; Serin, Lloyd, & Chadwick, 2019; Serin, Lloyd, & Hanby, 2010; Vose, Lowenkamp, Smith, & Cullen, 2009, etc.) has argued in favor of the inclusion of dynamic factors in risk assessment instruments. Conclusions from numerous studies including Douglas and Skeem (2005), Hanson and Harris (2000), and Kraemer, Kazdin, Offord, Kessler, Jensen, & Kupfer (1997) indicate that the initial assessment and subsequent monitoring of dynamic risk factors should be emphasized throughout treatment and supervision efforts, advancing that in order to be effective, risk assessment approaches must consider those factors that are changeable, and for which change has been shown to reduce risk of reoffending. Likewise, Brown, St. Amand, and Zamble (2009), Douglas and Skeem (2005), Lowenkamp, Johnson, Trevino and Serin (2016) and Vose, Lowenkamp, Smith, and Cullen (2009), support the inclusion of dynamic risk factors in assessment tools, maintaining that factors demonstrated to be changeable

should be a central focus for correctional staff, and for community supervision officers in particular. In other words, effective risk reduction begins with accurate risk (re)-assessment.

Risk, Dynamic Risk, and Protective Factors

Conceptualizations of risk. When correctional practitioners first began evaluating the risk of JIPs in a systematic way, the available assessment instruments were overwhelmingly comprised of static factors. As such, there was no opportunity to document intra-individual change of any kind barring re-arrest or new convictions. In the 1990's research started to emerge suggesting that monitoring changes in individuals' circumstances predictive of risk (i.e., dynamic risk factors; DRFs) was advantageous with respect to understanding and predicting reoffending behaviour, and interest in incorporating DRFs into case management practices increased. For instance, following their meta-analytic review, Gendreau and colleagues (1996) concluded that DRFs predicted general recidivism at least as well as static factors. Likewise, findings from Zamble and Quinsey's (1997) study, which aimed to investigate recidivism as a process and to evaluate the utility of dynamic factors in predicting release performance, provided initial support for utility of DRFs in the prognostication of general recidivism.

In the decades since, there has been a proliferation of research concerning the predictive accuracy and associated case management implications of DRFs. There are now numerous studies supporting the utility of dynamic factors in the prediction of specific types of recidivism among various subpopulations of JIPs (e.g., violence risk in forensic inpatients [see Wilson, Desmarais, Nicholls, Hart, & Brink, 2013]; and violence risk among high-risk JIPs with psychopathic traits [see Lewis, Olver, & Wong, 2012]).

For instance, findings from a recent study by Lowenkamp, Johnson, Trevino, and Serin (2016) indicated that key acute factors (i.e., anger/hostility, negative mood, and access to victims) predicted violent rearrest in a federal probation sample. When considered in conjunction with findings from other studies like Howard and Dixon's 2013 study, which identified relationships between changes in dynamic risk factors and risk of reoffending, there is mounting support for the overall importance of considering DRFs. Notably, Howard and Dixon's (2013) findings also demonstrated that reassessments could yield incremental improvements in predictive accuracy. Similarly, Brown, and colleagues (2009) established that predictive accuracy could be improved as a function of systematic reassessments of DRFs, although they also noted that the prediction models that performed best overall contained DRFs considered in conjunction with static factors.

That said, not all of the research on DRFs has been supportive. Caudy, Durso, and Taxman (2013) examined the incremental predictive validity of DRFs relative to static factors. While findings indicated that key DRFs (i.e., substance use, education/employment, associates, attitudes) did demonstrate statistically significant relationships with reoffending, the inclusion of these factors failed to yield substantive improvements in predictive accuracy. Instead, their findings suggested that criminal history, a static factor, remained the strongest predictor of recidivism. Hence, the inclusion of DRFs in risk assessment and prediction remains a somewhat contentious issue. Despite mounting support, evidence is far from unequivocal, and further research is required given the far-reaching implications that DRFs have for risk assessment and, by extension, for case management.

Dynamic risk. In view of the foregoing, it is important to clarify what exactly is meant by several of the key terms used repeatedly throughout corrections literature. As observed decades ago by Kraemer et al., (1997), there is a lack of consistency in how pervasive terms like *risk* and *risk factors* are used, leading to ambiguity in the communication of scientific results. Durrant (2016) nicely summarizes the assortment of ways that various researchers have used to theoretically conceptualize risk factors in recent decades, noting the overall amount of variability and inconsistent blend of descriptive and explanatory elements with concern. Despite having been identified as a problem twenty years ago, this irregularity, which is likely exacerbated by the lack of a unified theory, remains an issue.

One of the broadest definitions in common usage is that put forth by Kraemer et al. (1997), which defines a risk factor for reoffending in simple terms, as any measurable feature of an individual that both temporally precedes and has a statistically significant relationship with recidivism. Risk factors can include any demographic, situational, or intra-individual characteristics that are statistically correlated with recidivism, and which can be shown to have existed or occurred prior to the reoffence. Given the vagueness of this definition, it is unsurprising that confusion surrounding the conceptualization of risk factors exists.

Various classifications and subdivisions have been proposed to cope with the sheer multiplicity of risk factors. The highest-order and most well-known distinction is the separation of risk factors into static and dynamic factors. Static risk factors, often referred to as historical factors, are factors that are empirically related to reoffending, yet are unchanging, and thus cannot be addressed during interventions (Bonta, 1996; Zamble

& Quinsey, 1997). Conversely, dynamic risk factors are disposed to change, and can vary as a function of the passage of time and/or in response to intervention. By definition, changes in dynamic risk factors are related to an individual's likelihood of engaging in criminal behaviour (Andrews & Bonta, 2010). Perhaps as a function of their obvious relationship with reintegration efforts, dynamic risk factors are often equated with criminogenic needs (e.g., Andrews & Bonta, 2010). However, some (e.g., Yang & Mulvey, 2012) have argued that a statistical relationship with recidivism is insufficient, and that purportedly dynamic risk factors should also be able to help pinpoint when failure is imminent. The contention that dynamic risk variables should be able to assist in identifying immediate precursors to reoffending raises additional questions, including questions regarding when or how often assessments comprised of DRFs should be conducted and whether it can then be inferred that more proximal assessments aid in prediction. The view that dynamic factors should be assessed regularly is being espoused with increasing frequency in recent research (e.g., Lloyd, 2015; Serin et al., 2015; Skeem & Monahan, 2011).

In addition to the categorization of risk factors as either static or dynamic, Skeem and colleagues (Douglas & Skeem, 2005; Skeem & Mulvey, 2002) argue that to gain an accurate picture of an individual's overall level of risk, one must first understand the distinction between *risk status* and *risk state*. Risk status is defined as the degree of inter-individual risk determined almost exclusively by unchangeable static factors, whereas risk state is defined as intra-individual risk predicated largely on extant standing on DRFs (Douglas & Skeem, 2005). As such, risk status, which is largely fixed, is most useful with respect to determining which individuals are at greatest risk to reoffend, whereas risk

state provides information about changes in a given individual's likelihood of reoffending (Serin et al., 2016). In other words, risk status and risk state can be generally understood as the "who" and "when" in relation to recidivism.

Other distinctions within the larger umbrella categories of static risk factors and DRFs have also been proposed. For example, Kraemer and colleagues (1997) further divided static factors into inherently unchangeable factors, (e.g., age at first conviction; termed *fixed markers)* and those factors that can change, albeit not as the result of an intervention (e.g., age, number of prior convictions, termed *variable markers*). Likewise, Hanson and Harris (2000) saw a need for additional subcategories within DRFs, dividing them into stable and acute DRFs. *Stable* DRFs are risk factors that can change, but that are somewhat resistant to change. For instance, factors such as personal proclivities, entrenched patterns of behaviour, as well as skills and competencies (or lack thereof) can be considered stable DRFs because they are correlated with recidivism and can change, although the process of effectuating any meaningful change is often very time and effort intensive for both the JIP and the treatment or supervision provider (Hanson & Harris, 2000; Hanson, Harris, Scott, & Helmus, 2007). Conversely, *acute* DRFs change very rapidly, and are any external conditions, environmental or internal circumstances (e.g., intrapersonal stressors), or other events that have the potential to change quickly (i.e., over the course of days or even hours), and that are empirically associated with imminent reoffending (Hanson & Harris, 2000; Hanson et al., 2007; Serin et al., 2019). Importantly, this conceptualization of dynamic risk factors as stable or acute is used in the DRAOR.

As noted by Hanson and Harris (2000), it is essential to match the type of risk assessment being used to the type of factor under consideration. Stable DRFs (e.g.,

personality disorders, pervasive patterns of behaviour, etc.) can be used for long-term, and global risk assessments, as enduring changes in these areas can result in reductions in overall level of risk for recidivism (Hanson & Harris, 2000). That said, the value of stable DRFs rests predominately in their ability to gauge lasting changes such as the effectiveness of interventions and treatment programs, and the successfulness of reintegration efforts. In contrast, acute DRFs are most relevant for risk assessments with JIPs on community supervision. Given that they are highly transient, monitoring fluctuations in acute factors (e.g., substance use, negative moods) can alert supervision officers that likelihood of recidivism has increased, thus allowing them to alter supervision strategies proactively.

A corollary of the distinctions proposed above is the emergence of a three-tiered classification system of risk factors. Placement within the classification system is determined as a function of how quickly a given risk factor is liable to change in addition to the temporal proximity of the change in relation to the observed offending behaviour. Predictably, the three tiers of factors are static, stable, and acute. Static risk factors are the most temporally distal factors (read: farthest away from reoffending behaviour) and are also the most stable (read: unchangeable). Their primary utility rests in determining an individual's overall level of risk of recidivism, often referred to as risk status. Stable risk factors make up the middle tier. These factors are dynamic, and therefore must be able to change, but do so slowly, and often only in response to effortful intervention. As such, they represent appropriate rehabilitation targets. The final tier is occupied by acute risk factors. These factors change the most rapidly and are the most temporally proximal to the reoffending behaviour. Given that acute risk factors are essentially indicators of

imminent criminal behaviour, they are often conceived of as representing an individual's risk state and are monitored closely during the supervision process.

In view of the foregoing, it is particularly important to remember that, regardless of conceptualization, risk factors for reoffending should not be assumed to be the same as the causal processes underlying the initiation of offending. Durrant (2016) argues that risk factors (dynamic or otherwise), are simply features of individuals that have been found to correlate with reoffending. He contends that risk factors can most accurately be considered descriptive characteristics with limited explanatory value in and of themselves; while they may assist in the prediction of recidivism, they do not explain why offending began in the first place. Durrant (2016) asserts that the confusion surrounding this issue stems primarily from a collective failure to clearly distinguish between predicting recidivism and predicting the onset of offending. Relatedly, Durrant (2016) claims that part of the reason that risk factors perform so poorly with respect to offering meaningful explanations of offending (as opposed to reoffending) behaviour is that they have been identified as a result of research endeavors focused explicitly on predicting recidivism – the populations of interest in these studies were comprised of correctional populations with existing criminal histories. Thus, although the risk factors derived from research may serve to discriminate between JIPs who are more and less likely to reoffend, these same factors are likely insufficient to distinguish members of the community who are predisposed to engage in offending behaviour in the future from those who would abstain.

Clearly, there is a need for continued research about dynamic risk factors; not only does additional elucidation of these factors stand to advance correctional practice by

contributing to the existing literature surrounding risk assessment and accurate prediction, as well as offender management (e.g., initial classification, treatment/intervention, other rehabilitation efforts, and supervision practices), but also has the potential to contribute to the construction of meaningful theories of crime.

Protective factors. Recently, increased attention has been directed toward explicit consideration of JIPs' assets. Colloquially referred to as strengths, these protective or promotive factors denote positive characteristics or resources available to the individual that lessen the likelihood of reoffending (Farrington, 2003; Jones, Brown, Robinson, & Frey, 2015; Rogers, 2000). The concept of protective factors initially arose in response to concerns that correctional researchers and practitioners were becoming overly focused on the identification and assessment of risk factors among correctional populations while largely failing to consider the possible existence of positive attributes. Researchers such as Rogers (2000) reasoned that while assessing risk with the intent of preventing or minimizing negative outcomes had undeniable merit, the positive attributes of an individual and their situation (i.e., protective factors) were often disregarded. Rogers also contended that protective factors needed to be taken into consideration, given the mediating or moderating role these factors played in the risk of recidivism equation. Furthermore, recent research (Serin & Lloyd, 2009; Serin et al., 2010) affirms that further investigation regarding the impact of including strength factors in assessment instruments is necessary as it may contribute to a better understanding of the pathways underlying crime initiation and desistance. At present, the need for unbiased assessment practices is commonly accepted, and interest in this newer avenue of research has produced numerous insights – that said, it has (as was the case with DRFs) resulted in multiple

inconsistent conceptualizations of protective factors and accompanying terminological ambiguity.

How protective factors (PFs) should be defined is clearer from an empirical perspective than from a conceptual one (Polaschek, 2016). Essentially, from a purely empirical standpoint, a PF predicts decreased involvement in criminal behaviour (Farrington, 2007). This conceptualization, also espoused by Borum, Bartel, and Forth (2005), implies that protective factors can be identified empirically as any variable that evinces a negative correlation with offending behaviour. Alternatively, others, (e.g., Brook, Whiteman, Gordon, & Cohen, 1989; Hawkins, Catalano & Miller, 1992; Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Rutter, 1985; Webster, Martin, Brink, Nicholls, & Middleton 2004) have postulated that protective factors can be understood as existing at the opposing ends of a continuum with risk factors; thus, the same factor can function as either a vulnerability or an asset. Similarly, the absence of risk factors has been viewed as protective by some (e.g., Costa, Jessor, & Turbin, 1999). A notable distinction between the first and second conceptualizations described above is related to the concept of polarity. In the first conceptualization, PFs are understood as being unipolar (i.e., present to some degree versus absent), meaning that any relationship they demonstrated with criminal behaviour would be considered independent of any other factors (Lösel & Farrington, 2012; Polaschek, 2016). Conversely, in the latter conceptualization, risk and protective factors are operationalized as being bipolar (read mirror images of each other), and are therefore associated with a corresponding risk factor; whether a factor is viewed as protective or aggravating is a function of the direction in which the factor is present (Polaschek, 2016).

Many (e.g., Costa et al., 1999; Jessor, Van Den Bos, Vanderryn, Costa, & Turbin, 1995; Serin & Lloyd, 2009; Serin et al., 2016; etc.) have adopted the view that PFs are completely independent from risk factors, and therefore exist without the need for a corresponding opposite risk factor. This view is inherently more flexible, as it makes it possible for a JIP to exhibit multiple risk and protective factors concurrently. Consider for example a JIP with a substance use problem. Is the lack of a substance use problem protective? How exactly would the opposite of a substance use problem be represented on a bipolar continuum? That said, some PFs can have a linear relationship with a matched risk factor. Level of intelligence is a good example for illustrative purposes – low intelligence can increase risk of criminal behaviour, whereas high intelligence can potentially protect against it (Farrington, 2007; Polaschek, 2016).

How (or if) these putatively protective factors are related to risk factors remains an area of deliberation. Both Baird (2009) and Harris and Rice (2015) have criticized the bipolar conceptualization as redundant, affirming that it reduces PFs to being no more than positively worded risk factors. On the other hand, approaches that assume that PFs and risk factors are independent of one another (and not simply the inverse) have been critiqued as a function of correlations found between these purportedly unrelated factors. This is of particular import for the present project as this criticism has been applied to the DRAOR by de Vries Robbé, de Vogel, and Douglas (2013). The total score obtained when using the DRAOR (Serin, 2007) is obtained by summing risk factors and subtracting protective factors and can therefore be viewed as a risk/protection index. Thus, when studies involving the DRAOR (e.g., Yesberg & Polaschek, 2015) present moderate to strong correlations between the theoretically independent PF and risk factor

subscales, the assumption that these factors are truly independent is called into question (de Vries Robbé, de Vogel, & Douglas, 2013).

Increasingly, PFs are being perceived as variables that interact with risk factors to decrease the likelihood of engaging in criminal behaviour (e.g., Fougere & Daffern, 2011; Lodewijks, de Ruiter, & Doreleijers, 2010; Lösel & Farington, 2012; Ulrich & Coid, 2011. Although multiple frameworks have been proposed to explain the conceptual nature of the relationship between strengths-based factors and criminal behaviour, the present project espouses the framework described by Jones, Brown, Robinson, and Frey (2015). First, the framework presented by Jones and colleagues (2015) is intuitively appealing – it offers precise operational definitions for terms used inconsistently throughout the literature (i.e., strengths, protective factors, and promotive factors), and explains the mechanisms through which they are assumed to interact. Second, their framework aligns well with the aims of the present project. Specifically, the conceptualization of PFs espoused in Jones et al.'s (2015) framework is consistent with the one underlying the development of the DRAOR (Serin, 2007; Serin et al., 2007; Serin et al, 2016). Within this framework, the term *strength* is used to refer to any positive characteristic that could conceivably act as a safeguard (i.e., buffer) against an individual's risk for criminal behaviour. In accordance with Farrington's (2003) initial definition, Jones and colleagues (2015) define *promotive* factors as variables that demonstrate a negative relationship with criminal behaviour, irrespective of level of risk. For a variable to be considered a *protective* factor, Jones et al. (2015) theorize that the factor must influence an individual's risk for criminal behaviour differentially based on their overall level of risk; thus, genuine protective factors would be more salient for

higher risk JIPs, and correspondingly, would reduce risk of recidivism to a greater degree for high risk individuals versus low risk individuals.

The above conceptualization of strengths and PFs has important implications for release planning and desistance. Readers interested in a more comprehensive review of how the various ways in which protective factors are defined and conceptualized in relation to risk factors or about how protective and risk factors combine to influence recidivism, and/or how internal individual mechanisms like an individual's sense of agency, identity, and hope, as well as their overall attitudes toward crime and desistance may mediate the desistance process are encouraged to review some of the recent work by Polaschek (2016), and Serin and colleagues (e.g., Serin et al., 2015; Serin & Lloyd, 2009) respectively.

To review, notwithstanding considerable advances with respect to the field's understanding of protective factors and their integration into assessment instruments, many challenges still exist. Some researchers (e.g., Baird, 2009; Harris & Rice, 2015) remain unconvinced that protective factors are distinct from risk factors (i.e., not just reverse-worded). Moreover, Harris and Rice (2015) contend that increases in predictive accuracy attributed to protective factors found in some studies may not really be due to protective factors themselves, but rather simply be a function of the assessment touching on relevant factors that were otherwise not being addressed. Another concern is that voiced by Ward (2017), who contends that in the absence of a more complete understanding of the theoretical underpinnings of protective factors in relation to criminal behaviour, evaluations of the predictive utility of these alleged factors is premature. Although Ward's point should not be summarily dismissed as a more unified and

developed theory is certainly desirable, neither should it be accepted without question; the pathways to theoretical advancements are not often linear, and it would seem ill-advised to curtail promising avenues of research on these grounds. As emphasized by Jones et al. (2015), the inclusion of protective factors has been shown to yield improvements in predictive validity and have meaningful implications of terms of day-to-day practical utility.

Dynamic risk and protective factors in assessment tools. Advancements in theoretical understanding and mounting empirical support for dynamic risk and protective factors has been paralleled by the development of a number of assessment tools containing these variables. While these tools are similar insofar as they all incorporate dynamic risk and protective factors in some sense, they are nevertheless quite diverse with respect to the types and number of dynamic factors they incorporate, how they conceptualize risk and protective factors (e.g., the nature of the relationship between risk and protective factors), and the type of JIP population for which they were developed. For instance, contemporary measures that incorporate dynamic risk and protective factors include the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004), the Structured Assessment of PROtective Factors for violence risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009; 2012), the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006), the Inventory of Offender Risk, Needs, and Strengths (IORNS; Miller, 2006), the Service Planning Instrument (SPIn; Orbis Partners, 2003), and the DRAOR.[1]

[1]Readers interested in an in-depth treatment of the specific items comprising the other instruments listed above should refer directly to the measures or to some of the recent work by Serin and colleagues. For example, Serin, Chadwick, and Lloyd (2015) provide a thorough overview of the extant literature on the identification and measurement of dynamic risk and protective factors, as well as comparison of several

The DRAOR (Serin, 2007) is a structured professional judgment (SPJ) assessment tool designed for use with an adult, general JIP population, and can be considered a hybrid model, as it is predicated on a synthesis of contemporary risk assessment and desistance literature (Serin & Lloyd, 2009; Serin at al., 2010). It was specifically developed to assist probation and parole officers to assess and reassess their clients' changing needs and circumstances throughout the supervision process (Serin et. al., 2015; 2016), and, in line with the expressed objective of guiding case management decisions, is comprised entirely of dynamic items measuring both risk factors and protective factors. Since its inception in 2007, the DRAOR has been piloted and incorporated into routine practice in New Zealand, and in one jurisdiction in the United States. The majority of findings regarding the DRAOR thus far are encouraging; the DRAOR's strengths-based items have been found to demonstrate a promotive effect, and DRAOR scores have also provided incremental predictive validity above and beyond static risk (e.g., Chadwick, 2014; Hanby, 2013; Lloyd, 2015). The point of import here is that DRFs and PFs, such as those comprising the DRAOR belong in risk assessment tools.

Dynamic risk and protective factors in prediction. How dynamic factors are considered and what they can tell us as researchers varies as a function of the correctional context. Importantly, research on rehabilitation and desistance views dynamic risk and protective factors with an eye to how they contribute to a theoretical understanding of pathways to and from crime (i.e., as mechanism of change), whereas the risk assessment literature is instead focused on the predictive utility of these factors (Lloyd, 2015). While

---

contemporary assessment measures vis-à-vis the dynamic factors included, and Serin, Lloyd, & Chadwick (2019) offer an evaluation of the field's current level of conceptual understanding regarding the nature and utility of dynamic factors and discuss how these factors might be most effectually integrated into community supervision practices via inclusion in assessment tools.

there are still some (i.e., Gottfredson & Moriarty, 2006) who would argue that risk assessment should be conducted purely on the basis of static factors given that they have been proven to be more consistently powerful predictors and are more easily defined and measured than dynamic factors, this interpretation is increasingly rare. Nonetheless, studies examining the predictive abilities of truly dynamic risk and protective factors remain scarce, and those with strong methodologies are scarcer still. Evidently, the suitability of an analytical approach is largely contingent on the research objectives.

Chapter 2: The Influence of Gender, Race, and Intersectionality on Dynamic Risk Assessment

It is difficult to quantify the extent to which marginalized groups such as women and racial or ethnic minority groups are involved in the justice system.[2] Official crime statistics (e.g., the FBI's Uniform Crime Reports) do not disaggregate findings according to gender, race, and ethnicity, thus, it is often necessary to rely on unofficial sources such as unpublished crime reports (Chilton & Datesman, 1987; Steffensmeier & Allan, 1988), supplementary reports (Cooper & Smith, 2011), or self-report research (Cernkovich & Giordano, 1979).

For example, Cooper and Smith (2011) used supplementary homicide reports to examine arrest rates for Black and White men and women in an effort to explore the Black women/White men crime convergence hypothesis. Intriguingly, despite a lack of strong empirical evidence, many early scholars (e.g. Adler, 1975; Pollack, 1950; Smith & Visher, 1980; Sutherland & Cressey, 1978) speculated that Black women and White men

[2] The term 'race' is used to reflect both race and ethnicity throughout this dissertation to simplify discussion of these topics and reduce confusion. See Chapter 4 (description of sample) for the rationale underlying this choice of terminology.

committed crimes at a similar rate. Although Cooper and Smith (2011) did find some evidence to support this hypothesis, the observed differences were much smaller than what earlier scholars would have predicted. The study, which investigated homicide rates in the USA, found that although Black women were responsible for homicides at a greater rate than White women (10 per 100,000 versus 2.2 per 100,000), that the homicide rates of Black women and White men were still highly discrepant. In fact, the homicide rate for White men (20.4 per 100,000) was still more than double that of Black women. Findings such as these highlight the need to eschew popular theoretical perspectives unless they are backed by substantive empirical evidence. The following section is intended to provide some essential context regarding risk assessment with two oft ignored correctional subpopulations, namely JI women, and JIPs belonging to historically marginalized groups. How these identities interact (i.e., intersectional theory) to dramatically increase vulnerability to and within the criminal justice system is also discussed.

Regrettably, gender and race remain contentious and poorly understood issues within the criminal justice system. While mainstream correctional scholars (e.g., Andrews & Bonta, 2010) advance that dynamic factors and contemporary assessment tools largely transcend gender and race, others contend that this is not the case, and that, at the very least, different variables may be more or less salient for both females (e.g., dysfunctional relationships; Brown & Motiuk, 2008; Benda, 2005; Van Voorhis, Wright, Salisbury, & Bauman, 2010; Van Voorhis, 2012) and historically marginalized groups (e.g., substance use; Boer, Couture, Geddes, & Ritchie, 2003; Trevethan, Moore, & Rastin, 2002). By extension, these authors contend that such factors should therefore

inform treatment priorities. The following section addresses the application of dynamic risk and protective factors to JI women and JI individuals belonging to historically marginalized groups .

Race

The overrepresentation of historically marginalized groups. When the composition of the general population in North America in general and the United States in particular is compared to the composition of its correctional population, it becomes evident that racial historically marginalized groups are vastly overrepresented. Carson (2015) reported that in excess of one third (37%) of the prison population in the United States was categorized as Black, with 32% and 22% of the JIPs being categorized as White, and Hispanic, respectively. In comparison, data retrieved from the 2014 United States Census estimated that Black individuals only represented approximately 13.1% of the overall population, whereas Whites comprised almost two thirds (62.1%) and Hispanics accounted for 17.4%. This disproportionality is even more evident when considering only JI women. According to Glaze (2011) White women are incarcerated at a rate of roughly 91 per 100,000 where are the incarceration rates for Hispanic women and Black women are 1.5 times higher and nearly three times higher (133 per 100,000 and 260 per 100,000) respectively. Somewhat predictably, in spite of the fact that non-White women are overrepresented in the criminal justice system, the majority of extant female-oriented research has been largely concerned with the majority demographic (i.e., White women).

At this point, it should be noted that the composition of the population in Iowa differs appreciably from the composition of the overall American population. Current

census data estimates that in excess of ninety percent (91.4%) of the population of Iowa is White, whereas Blacks and Hispanics account for an estimated 3.7% and 5.8% respectively. That said, historically marginalized groups, and those categorized as Black in particular, are still dramatically overrepresented within the Iowa Department of Corrections (IDOC, 2017). For example, JIPs classified as White represent 65% of the IDOC prison population, and 76% of the community-based corrections (CBC) population (IDOC, 2017). In comparison, Blacks account for 25% and 16% of the prison and CBC populations respectively, and Hispanics for 7% and 5% (IDOC, 2017). Clearly, racial disproportionality is a pervasive issue for the IDOC. Unfortunately, specific underlying causes of the overrepresentation of individuals belonging to historically marginalized groups, and relatedly, what to do about them, remains unclear.

Concerns about the role of risk assessments in corrections. As noted by Flores, Bechtel, and Lowenkamp (2016) "one of the most common concerns that arises in the discourse on the use of risk assessment in correctional and sentencing contexts is racial bias" (p. 39). The prospect of racial bias in risk assessment has been a concern for some time in the United States (Henderson, Tanana, Wyatt Bourgeois, & Adams, 2015), and this issue is further complicated by ethical considerations and the sheer size of the correctional population, which is accompanied by substantial financial and human costs (Flores et al., 2016; Skeem, 2013; Skeem & Lowenkamp, 2016).

The current state of affairs on this topic is nicely illustrated by a recent study by Angwin, Larson Mattu, and Kirchner (2016), and the subsequent rejoinder by Flores et al. (2016). Angwin and colleagues (2016) investigated racial bias in the Northpointe COMPAS (1998), a commonly used actuarial risk assessment instrument (ARAI), and

published an article implying that it, and likely all other ARAIs were biased towards historically marginalized groups. Further, they concluded that ARAIs are therefore inappropriate for use in sentencing and other correctional decision-making contexts. Flores et al., (2016) directly challenged Angwin and colleagues' findings, citing methodological shortcomings, and the fact that their findings contradicted those of several other comprehensive studies that had demonstrated that ARAIs could be used to predict risk, free of racial and/or gender bias. Still, other researchers (i.e., Hannah-Moffat, 2010), have expressed concern that the application of risk assessment technology in sentencing and other correctional settings will exacerbate existing bias against marginalized groups, and some (e.g., Holder, 2014; Starr, 2014; Van Eijk, 2017) have condemned the use of ARAIs even more strongly, asserting that reliance on such instruments is likely unconstitutional and discriminatory. However, extant evidence seems to indicate that, if judiciously applied, risk assessments can have a positive impact on the correctional system; even though well-validated risk assessment instruments cannot (and do not claim to) predict criminal recidivism at the individual level with 100% accuracy, they can demonstrably improve prediction above and beyond unstructured professional judgement (Skeem & Monahan, 2011).

A second issue put forward in conversations about racial bias is the extent to which utilization of risk assessment tools might aggravate existing biases against marginalized groups. Proponents of this broader perspective (e.g., Hannah-Moffat, 2010; Harcourt, 2015; Holder, 2014; Silver & Miller, 2002; Starr, 2014; Van Eijk, 2017) worry that risk assessment tools contribute to disparities in sentencing, and lead to the individualization of social problems, like racial discrimination and socioeconomic bias.

For instance, Harcourt (2015) contends that the criminal history variable, included in the vast majority of assessment tools, is a proxy for being Black. In a similar, yet distinct vein, Starr (2014) claims that it is illegitimate to assess an individual's risk of recidivism based on instruments that assess factors like marital history, employment status, education, and financial background (although she does not take issue with risk assessments' reliance on the criminal history variable, which she feels is justified). Essentially, her concern is that reliance on risk assessment instruments will result in longer sentences for marginalized groups based on factors and circumstances that are largely beyond their control.

Importantly, well-validated assessment tools generally include empirically and theoretically grounded predictors of criminal behaviour, and do not include race (Skeem, 2013). For example, findings from an earlier study (Andrews, 2012) corroborate the ability of well-validated risk assessment tools to predict recidivism above and beyond the effects of race, poverty, and gender. Manifestly, this contradicts the view that variables that have consistently been shown to be robust predictors of recidivism (e.g., patterns of antisocial behaviour, impulsive or aggressive traits, criminal associates, etc.) are nothing more than proxies of social disadvantage (Skeem, 2013). Regardless, continued research is advisable. As noted by Durrant (2016), given that predictive utility is often the foremost consideration in the development of risk assessment instruments, it is concerning that many of the constructs, mechanisms, and underlying causes that they symbolize remain poorly understood.

Current evidence. Given the racial disproportionality in the correctional system, concerns surrounding racial bias are legitimate. Also, while most of the criticisms

concerning racial bias have focused on the front-end of the process (e.g., the inclusion of risk assessment in sentencing), the same concerns are also germane to back-end decision-making (e.g., determinations of whether an individual should be released, conditions of probation/parole, etc.; Skeem & Lowenkamp, 2016). When attempting to apportion the causes of this disproportionality, it is important to acknowledge that disparities in per capita incarceration rates are likely to reflect an interaction between several different factors, namely: (a) racial differences in the prevalence, incidence, and nature of criminal behaviour; (b) actual bias or discrimination; and (c) the unequal influence of apparently race-neutral policies (e.g., in sentencing, case-processing, etc.) throughout the criminal justice system (Frase, 2013).

So, what evidence is there to substantiate or discredit concerns about extant risk assessment instruments and racial bias specifically with respect to predictive accuracy? As stated by Skeem and Lowenkamp (2016), "data may be more helpful than rhetoric if the goal is to improve sentencing and correctional practices" (p. 684); to determine if a test is valid, reliable empirical evidence is necessary.

Accordingly, Skeem and Lowenkamp (2016) used the Post Conviction Risk Assessment (PCRA; Johnson, Lowenkamp, VanBenschoten, & Robinson, 2011), a risk assessment instrument designed and validated for use in guiding decisions about service and supervision needs at intake of community supervision (not during sentencing), to empirically examine the relationships among race, risk assessment, and recidivism in a sample of Black and White JIPs in federal custody. First, the authors examined the predictive fairness of the PCRA (i.e., whether the prediction of rearrest varied as a function of the JIP being Black or White). Area under the curve analyses demonstrated

that PCRA total scores strongly predicted both non-violent and violent rearrest (*AUC* = .71, .72 and .74, .75) for Black and White JIPs respectively, and subsequent regression analyses yielded a non-significant interaction term, suggesting that race did not moderate the predictive ability of the PCRA. Second, Skeem and Lowenkamp (2016) assessed the extent to which the Black and White JIPs obtained different PCRA total scores. Consistent with their hypotheses, Black JIPs did score slightly higher on the PCRA than White JIPs (*M* = 7.37, *SD* = 3.25 vs *M* = 6.23, *SD* = 3.38), and the effect of race was found to be *d* =.34 (which equates to 13.5 percent non-overlap and 86.5 percent overlap) in PCRA scores. Next, the authors examined which factors contributed the most and the least to mean score differences; findings indicated that criminal history accounted for roughly two thirds (66%, *d* = .34) of the variation in scores, whereas variable (read dynamic) factors like substance use, attitudes, and social factors contributed negligibly to mean score differences between Black and White JIPs. Last, Skeem and Lowenkamp (2016) determined that the criminal history factor, which they note is already embedded in sentencing guidelines, was not a proxy for race, but instead mediated the relationship between race and future arrest.

Similarly, Flores et al. (2016) failed to find evidence of predictive bias on account of race in their investigation of the Northpointe COMPAS. Flores and colleagues' (2016) findings demonstrated that while Black defendants did in fact have a higher base rate of failure than White defendants (52% versus 39%), that these differences are not indicative of bias, but instead simply described the behaviour of defendants in the criminal justice system. Relatedly, the authors also found that failure rates increased monotonically within the risk categorizations of the COMPAS, indicating that while the base rate of

recidivism was higher for Black defendants, that general failure rates for both White and Black defendants increased proportionally across low, medium, and high-risk categorizations. Next, Flores et al. (2016) used a series of receiver operating characteristics (ROC) analyses to examine the predictive accuracy of the COMPAS for all defendants, as well as across defendants' race. Results indicated that the COMPAS strongly predicted recidivism for all defendants (*AUC* = .71), and that predictive accuracy was likewise high for both Black and White defendants, with *AUC* values of .70 and .69 respectively. Importantly, the slope and intercept of the relationship between COMPAS scores and general recidivism was comparable for Black and White defendants. In other words, the interaction term between race and COMPAS scores was not significant and did not improve the prediction model; a given score on the COMPAS is indicative of roughly the same likelihood of general recidivism, regardless of defendant race. Similar analyses focused on predicting violent recidivism also failed to find evidence of bias (Flores et al., 2016). Finally, logistic regression analyses estimating the form of the relationship between COMPAS scores and any outcome (i.e., any arrest and violent arrest) again revealed no differences in slope or intercept and produced a non-significant interaction term. In sum, all examinations of the COMPAS found that it predicted consistently well, and equitably for both Black and White defendants.

Summary and conclusions. Taken as a whole, the extant literature seems to suggest that there is less cause for alarm than some (e.g., Angwin et al., 2016; Holder, 2014; Starr, 2014) would claim. Studies such as Flores et al. (2016) and Skeem and Lowenkamp (2016) demonstrate that risk assessment instruments can make significant, positive, contributions in correctional settings, and specifically, that they can be

employed to detect legitimate differences in base rates by race while predicting outcomes in an unbiased way. Ultimately, historically marginalized groups stand to benefit from additional clarification of the predictive abilities of contemporary assessment tools, and the application of such tools may actually help reduce some of the disproportionality (Flores et al., 2016). Many (e.g., Flores et al., 2016; Olver et al., 2013; Skeem & Lowenkamp, 2016) have acknowledged the desirability of approaches such as factor analysis to test for invariance, and as such, the present project is well-positioned to make a contribution in this area.

Gender

Responding to the growth of the female correctional population. Justice involved women are widely recognized as one of the fastest growing correctional populations (Blanchette & Brown, 2006; Greiner, Law, & Brown, 2014; Guerino, Harrison, & Sabol, 2011; Public Safety Canada, 2016, etc.), and have therefore received increasing attention from correctional researchers and policy-makers alike. The reality is that the majority of contemporary risk assessment tools were developed and validated using samples comprised predominantly of JI men (Chesney-Lind, 1997). Brown, Serin, Forth, Bennell, Nunes, and Pozzulo (2017) use the term *gender-neutral* to describe such assessment tools, which were developed based on gender-neutral theories of crimes, tested on largely male samples, and subsequently used with women. As noted by Brown and colleagues (2017), critics of the gender-neutral perspective argue that describing an assessment tool as “neutral” is analogous to describing it as “inherently male”, and that such tools are biased against females, whereas defenders of the perspective counter that it simply means that it is equally suitable for use with both men and women. Greiner, Law

and Brown (2014) emphasize the need for continued female-specific correctional research, as historically, most primary studies and large-scale meta-analyses have focused on JI men. Encouragingly, recent years have seen a proliferation of women-centered research (e.g., Bloom, Owen, & Covington, 2003; Chesney-Lind & Pasko, 2013; Holsinger, 2000; Odgers & Moretti, 2002; Van Voorhis, 2012; Van Voorhis, Bauman & Bruschett, 2013), and corresponding advancements in the areas gender-specific risk assessment and treatment programming have taken place.

Female-perpetrated crime. With respect to criminal behaviour, males and females differ in several important ways. First and foremost, men markedly outnumber females in correctional populations, and are also much more likely to be involved in violent crime (e.g., Blanchette & Brown, 2006; Brown et al., 2017; Public Safety Canada, 2016, etc.). However, while males consistently outnumber females across all types of crime, the gender gap does vary as a function of type, and sub-type of crime (Brown et al., 2017). For instance, in Canada, if we consider the broader offence category of property crimes, males are responsible for nearly 90% of break-and-enters and motor vehicle thefts, whereas females accounted for roughly two thirds of general theft, and fraud-related crimes (Public Safety Canada, 2016; Statistics Canada, 2016).

Beyond basic differences in type and number of crimes committed, females are also hypothesized to engage in criminal behaviour (even similar criminal behaviours) for different reasons than their male counterparts (e.g., Campbell, 2002; Greenfeld & Snell, 1999). For example, numerous authors have investigated gender differences in homicide offences (e.g., Campbell, 1984; Daly & Wilson, 1988; Greenfeld & Snell, 1999; Kruttschnitt, 2001; Owen, 2001), and non-violent crimes (e.g., Belknap 2014; Campbell,

2002; Carlen, 1988; Chesney-Lind, 1986). Taken as a whole, these studies indicate that there are both differences and similarities in terms of (a) the types of crimes committed by males and females, and (b) their underlying motivations and circumstances.

Factors associated with female offending. Increasingly, correctional agencies are embracing the idea that JI women do differ from their male counterparts. While the precise nature and magnitude of these differences has yet to be determined, empirical evidence suggesting that certain risk and need factors may be more salient for JI women (e.g., substance use, mental health and self-esteem factors, etc.) is growing, as is the recognition that their treatment needs differ somewhat from those of JI men. These paradigm shifts are underscored by a significant (if still nascent) trend in the literature; the gender-difference hypothesis (Hyde, 2005) is being tested empirically with greater and greater frequency, as researchers have adopted methodological approaches involving mixed-gender samples allowing for explicit comparisons (Brown, 2017).

Initial reviews (e.g., Byrne & Howells, 2002; Sorbello, Eccleston, Ward, & Jones, 2002) concluded that risk factors were largely similar across genders. Likewise, findings from several earlier meta-analyses investigating factors associated with criminal behaviour in females (e.g., Green & Campbell, 2006; Hubbard & Pratt, 2002; Simourd & Andrews, 1994) suggested that the Central Eight (i.e., criminal history, antisocial personality, associates, and cognitions, substance use, family/marital relationships, school/work, and social/recreational activities; Andrews & Bonta, 2010) were gender-invariant. Dowden and Andrew's (1999) meta-analysis of female-specific treatment outcomes (revisited in 2005 by Dowden) yielded concurrent findings. Correspondingly, meta-analytic validation studies of risk assessment tools grounded in the Central Eight

such as the LS/CMI (Andrews, et al., 2004) have demonstrated that they are valid for women.

However, advocates for the gender-difference hypothesis such as Shaw and Hannah-Moffat (2004) and Van Voorhis (2012) have condemned findings from meta-analyses for their failure to take small-scale studies and qualitative research into consideration, and for directly comparing results between genders only infrequently. Andrews et al. (2012) addressed this second criticism, examining gender differences in the predictive ability of both the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta & Wormith, 2004) and the YLS/CMI (the youth version). Results indicated that all dynamic domains significantly predicted recidivism in both males and females. Noteworthy, substance use demonstrated a stronger association with recidivism for females than males. Concomitantly, other researchers have detected gender differences in the predictive validity of specific dynamic factors when using single-occasion measurement strategies. For example, Manchak, Skeem, Douglas, and Siranosian (2009) found that while overall Level of Service Inventory-revised (LSI-R; Andrews & Bonta, 1995) scores were equally predictive for both men and women in their sample of violent JIPs, both financial and substance use domains were strong predictors of recidivism for men, whereas only the financial domain was predictive for women. Similarly, when van der Knaap, Alderda, Oosterveld, and Born (2012) compared the predictive ability of various risk domains across gender, they noted several important differences; problems with accommodation, education and employment, and peer relationships correlated more strongly with recidivism among men, and emotional problems were more salient for women. Collectively, these findings suggest that, when

measured at a single time-point, some putatively gender-neutral dynamic risk factors may be moderated by gender (Greiner, et al., 2014).

Static criminal history variables (e.g., age at onset, number of crimes/frequency of criminal behaviour, etc.) have consistently been shown to be among the strongest predictors of reoffending, regardless of gender (Andrews & Bonta, 2010; Andrews et al., 2012; Gendreau, Goggin, & Smith, 2002; Green & Campbell, 2006; etc.). Conversely, when we consider individual dynamic risk factors, findings are more mixed. Findings generally suggest that antisocial personality/impulsivity and antisocial attitude variables predict recidivism in both male and female JIPs, although some incongruities have been detected (see Blanchette & Brown, 2006; Coid, Yang, Ullrich, Roberts, & Hare, 2009; Edens, Campbell, & Weir, 2007). For example, Stockdale, Olver and Wong (2010), found evidence for the predictive ability of the Psychopathy Checklist (PCL; Hare & Frazelle, 1980) measure for younger (adolescent), but not older (adult) JI women. Next, evidence that has emerged pertaining to mental health, self-esteem, and self-efficacy suggests that these individual risk factors may be more germane to JI women. Low self-esteem and low-self efficacy predicted criminal recidivism in a sample of JI women (Van Voorhis et al., 2010), as did depression and anxiety (Van Voorhis, Salisbury, Wright, & Bauman, 2008). Research has also found that childhood maltreatment is predictive of criminal behaviour in women (Andrews, Bonta, Wormith, Guzzo, & Brews, 2008; Lowenkamp, Holsinger, & Latessa, 2001). However, whether childhood maltreatment is a female-salient predictor of recidivism (i.e., predicts more strongly for women) remains unclear. Several studies (e.g., Green & Campbell, 2006; Hubbard & Pratt, 2002; Van Voorhis et al., 2008) support the notion of female-salience for this variable, while others

(e.g., Andrews et al., 2008; Daigle, Cullen, & Wright, 2007; Lowenkamp, et al., 2001) do not. Similarly, there is some evidence that relationships with criminal peers might be especially predictive of criminal recidivism among females. While having criminal friends is unequivocally an important risk factor regardless of gender (e.g., Brown & Motiuk, 2008; Hubbard & Pratt, 2002), Hipwell and Loeber (2006) suggest that criminal friends may influence females even more strongly as a function of the high degree of intimacy common in close friendships among females. Dysfunctional relationships between women and intimate, criminal partners have also been found to be problematic for both adult (Benda, 2005; Bloom et al., 2003), and adolescent women (Hipwell & Loeber, 2006). Finally, while lack of education and unstable employment are risk factors for both genders (Andrews & Bonta, 2010; Bloom et al., 2003), there is some evidence that improvements in skills related to maintaining consistent employment were related to reductions in reoffending among women, but not among men (Brown & Motiuk, 2008).

With respect to protective factors, it is worth mentioning that simply being female is considered a protective factor in and of itself (Moffitt & Caspi, 2001). To date, few studies have specifically investigated the differential impact of putative protective factors according to gender. One study that has (Jones et al., 2015) revealed that protective factors were equally prevalent in men and women, and that protective factors reduced risk of recidivism in both high- and low-risk JIPs, although the effect was more pronounced in the high-risk group.

Criticisms of gender-neutral instruments. One of the main criticisms of ostensibly gender-neutral instruments is that they fail to adequately consider the extent to which the gendered pathways (e.g., crime as a survival strategy [see Bloom et al., 2003;

Chesney-Lind & Rodriguez, 1983, Daly, 1992, 1994; or Simkins & Katz, 2002, for more information] or relational, childhood victimization, and social/human capital pathway models [see Salisbury & Van Voorhis, 2009]) impact the validity of risk assessment tools. Also, detractors of gender-neutral tools argue that important female-specific risk/needs factors (i.e., financial stress, internalizing mental health symptomology, history of victimization, dysfunctional relationships, and poor self-efficacy) are conspicuously absent (Van Voorhis, 2012). In sum, although findings concerning potential gender-based difference in the salience and predictive abilities of various dynamic risk and protective factors are still somewhat sparse, there is nonetheless sufficient evidence to warrant further research into gender-responsive assessment tools.

To address the criticisms outlined above, Van Voorhis and colleagues (2008, 2010) developed a female-specific assessment tool, the Women's Risk Need Assessment (WRNA), which was subsequently revised (the Revised WRNA). Briefly, the WRNA assesses gender-informed risks, needs, and strengths factors, many of which are distinct from those included in supposedly gender-neutral assessments (e.g., depression, housing safety, abuse, self-efficacy, etc.; Van Voorhis, Bauman, & Bruschett, 2013). Furthermore, shorter versions of the tool have been derived for use in pre-trial settings (Gehring & Van Voorhis, 2014), and to supplement existing risk assessment tools like the LS/CMI. Initial validation studies of the WRNA have yielded encouraging, if somewhat difficult to interpret results. Van Voorhis, Bauman, and Bruschett (2013) present numerous tables of bivariate correlations and *AUC*s that vary as a function of sample (e.g., Iowa, Minnesota), outcome (e.g., arrest, conviction, technical violation, any failure, etc.), and measure (e.g., WRNA standalone, LSI-R + WRNA). Overall, the *AUC*s were

relatively small according to Rice and Harris's (2005) conventions, yet significant (ranging from the .50s to the .70s; presented without confidence intervals), and findings would also appear to suggest that the gender-responsive WRNA was able to incrementally improve prediction above the LSI-R.

Likewise, Canadian researchers Orbis Partners have recently developed two new risk/needs measures, one for youth, and the other for adult women. Validation research describing the psychometric properties of the first, the Youth Assessment Screening Inventory for Girls (YASI-G; Orbis Partners, 2007b), is limited to date. Scott (2017) recently conducted a validation study of the original YASI, and a gender-informed version, the YASI-GI. Scott's (2017) findings demonstrated that the YASI-GI predicted general and violent recidivism with moderate predictive accuracy (*AUC*s .63 - .71), which is consistent with earlier work conducted by Orbis Partners (2007b), the scale developers, on the YASI. However, although Scott did identify some domain areas (e.g., attitudes, peers) predictive of violent recidivism among females, she was not able to reach definitive conclusions regarding gender neutrality, gender saliency, and gender specificity. Regardless, initial findings are promising, and Scott has identified avenues for continued research. The second measure, the Service Planning Instrument (SPin; Orbis Partners, 2007a), has undergone initial validation, and has yielded promising validity indices with samples of adult women (Jones et al., 2015).

Current evidence. A recent report by Brown (2017) provides a useful summary of the recent body of research pertaining to women's risk and need factors. Following a thorough review of relevant databases (i.e., PsychInfo, PsychArticles, Criminal Justice Abstracts, and Scopus) Brown (2017) identified and synopsized studies that directly

examined the hypothesis that JI women differ from JI men, emphasizing key findings related to the gender-neutrality, salience, and/or specificity of the risk factors investigated. The term *gender-neutral* was used to describe factors found to be equally predictive of recidivism in both men and women, whereas *gender-salient* was used to describe risk factors that predict recidivism in both genders, albeit more strongly for one gender when compared to the other (Brown, 2017). Finally, the term *gender-specific* was used to designate any risk factor that predicted recidivism exclusively in one gender. Brown's (2017) report suggests that there is currently more evidence for gender-neutrality over gender-salience/specificity with respect to risk factors. Nonetheless, her review of the literature also yielded several more specific, and important conclusions.

First, sufficient evidence has now accrued to demonstrate that the overarching domains of substance use and personal/emotional factors are women-salient predictors of recidivism. For instance, findings from Olver and colleagues' (2014) comprehensive and methodologically rigorous meta-analysis examining the predictive validity of all version of the Level of Service tools (youth-, adult- and region-specific versions) demonstrated that both substance use, and personal/emotional factors are female-salient, regardless of age. Similarly, numerous studies have found associations between substance use and recidivism in women (e.g., Cimino, Mendoza, Thielman, Shively, & Kunz, 2015; Gobeil, Blanchette, & Stewart, 2016; Kopak, Proctor, & Hoffman, 2015), and others have identified substance use problems as being stronger predictors of recidivism in women than in men (Chang, Larsson, Lichtenstein & Fazel, 2015). Moreover, studies such as Brown and Motiuk (2008) have uncovered nuanced gender differences and underscore the importance of delving deeper into widely accepted higher-order constructs like the

Central Eight (Andrews & Bonta, 2010). For example, with respect to substance use, their results indicated that age of onset and engaging in social drinking were particularly salient risk factors for males, but not for females. Conversely, for females, substance use as a means to cope with stress or that was severe enough to impair daily functioning was found to be especially predictive of criminal behaviour. Findings from another study indicated that alcohol and drug problems were predictive of rearrests among the male contingent of the matched sample, but not for the female contingent (McCoy & Miller, 2013). More research is required to determine if findings like these will replicate, and to better understand the implications that nuances such as these have in terms of risk assessment and treatment.

Second, Brown (2017) notes that much of the evidence for the above conclusion is born of studies centered on the LSI instruments. Of those reviewed in Brown's report, many of the primary studies assessing the predictive accuracy of dynamic factors for JI women do so using the overall scores and/or subdomains of the LSI (e.g., Andrews et al., 2012; Geraghty & Woodhams, 2015; Olver et al., 2014). Replication with other measures is desirable.

Third, there appears to be some equivocality with respect to gender-neutrality even within some of the global risk/needs domains generally recognized as being equally predictive for both males and females. Although evidence generally indicates that the broader domains of criminal peers, criminal attitudes, employment, marital/family functioning, and community functioning are gender neutral, findings from several studies suggest that there may in fact be nuanced gender differences. Results from Benda's (2005) analysis of criminal recidivism in a matched sample of male and female boot

camp graduates revealed several noteworthy gender differences; abuse, adverse or negative emotional states, and certain relational variables like having a criminal partner were female-salient predictors, whereas aggressive feelings, criminal peers, and education/employment variables were male-salient risk factors. Somewhat intriguingly, findings also indicated that while drug use use appeared to be a gender-neutral predictor, alcohol use predicted recidivism more strongly in men (Benda, 2005). Relatedly, others (i.e., Yang, Knight, Joe, Rowan, Lehman, & Flynn, 2015) found evidence of an interaction effect between gender and self-esteem, decision making confidence, and peer support. For males, high levels of self-esteem predicted re-arrest, whereas the opposite was true for females. Moreover, despite decreasing the likelihood of recidivism among males, peer support and decision-making confidence had no effect for females. As noted by Brown (2017), more research is necessary to see if these findings can be replicated.

Fourth, despite promising initial findings (e.g., Benda, 2005, Van Voorhis 2013), there is a lack of research explicitly investigating the true nature (i.e., female-saliency versus female-specificity) of generally acknowledged gender-responsive risk/need domains. Given that women consistently score higher on certain risk/needs domains (e.g., abuse, anxiety/depression, dysfunctional relationships, criminal intimate partners, parental factors, and housing concerns), the debate no longer concerns whether these factors merit consideration, but is instead centered on how and when they should be integrated (e.g., directly incorporated into risk assessment tools or treated as responsivity factors in programming; Brown, 2017).

Finally, current literature provides insufficient grounds to make definitive conclusions about the gender-specificity of any of risk/needs factors examined to date –

for males or females. Considered as a whole, the research supports the notion of gender-neutrality over gender-specificity or even gender-salience insofar as risk/needs factors are concerned. Important gender-differences may exist and may be uncovered in the future, but more research is necessary to clarify some of the nuanced gender-based findings that are beginning to emerge.

The Intersectionality Paradigm

The intersectionality paradigm is a perspective that has been championed by feminist criminologists and civil rights advocates for decades (Baca Zinn & Thornton, 1996; Crenshaw, 1989, 1993). This perspective contends that identities such as race, gender, and social class combine to multiplicatively worsen an individual's experience with the criminal justice system. This paradigm is centered around the perception that marginalized individuals (i.e., non-White, economically disadvantaged women) are most vulnerable to oppression, and that this oppression manifests as systemic discrimination in the criminal justice system (Brown et al., 2014). Said differently, the intersectionality paradigm stipulates that a marginalized individual such as a Black or Hispanic woman would have a significantly different experience with the criminal justice system than would a White man or woman. The following section provides a brief overview of the theory underlying the intersectionality paradigm and discusses the current application of mainstream correctional assessment and treatment practices to racially diverse and marginalized people.[3]

Theoretical origins and basis of intersectionality theory. The 1970s saw the organization of the first race-specific feminist groups in America. Examples include a

[3] Intersectionality issues are also salient for men belonging to racial and ethnic minorities. However, given the aims of this dissertation, more emphasis is given to theory and research on girls and women.

Mexican women's group (the Chicana group), the Asian Sisters, and the Women of All Red Nation (a Native American women's group; Campbell, 2002b). The National Black Feminist Organization (NBFO) was founded in 1973 in New York and produced a splinter chapter who established themselves in Boston. This splinter chapter of the NBFO is credited with writing the first black manifesto (the Combahee River Collective) and with providing the foundation for contemporary intersectionality paradigms (Baca, Zinn & Thornton Dill, 1996; Cole, 1999; Guy-Sheftall, 1995; Thompson, 2002). For more information on the evolution of the Collective, see Brown et al., (2014) and Guy-Sheftall (1995). The coining of the actual term *intersectionality* is credited to Crenshaw (1989, 1993), a legal scholar and critical race theorist. In her seminal work, Crenshaw identified four ways in which Black women plaintiffs in the criminal justice system experience discrimination: (a) in similar ways to White women; (b) in similar ways to Black men; (c) in ways which discriminate on the basis on gender and race in a cumulative manner; and (d) in a simultaneous, multiplicative way. Thus, Crenshaw argued that to fully understand the experience of a Black woman plaintiff, the influences of gender and race must be considered in an interactive (rather than additive) manner.

Related paradigms include multiracial feminism (Baca Zinn & Thornton Dill, 1996) and the matrix of domination (Collins, 2001), both of which also draw attention to the interlocking and multiplicative ways that facets of individual identity such as gender, race, class, sexuality, nationality, age, etc. interact to influence the extent of oppression experienced by a given woman. These three perspectives highlight the importance of considering all possible intersections of structural systems that converge to determine whether an individual is in a position of power and privilege, or marginalization and

discrimination. Notably, those who regard themselves as multiracial feminists tend to view race as the predominant cause for oppression whereas feminists who prefer the term intersectionality instead afford preeminent causal status to gender (Brown et al., 2014). Encouragingly, advances in research methodology now permit for more rigorous investigations of the interactions between these variables.

Explaining female-perpetrated crime. Historically, research about criminal behaviour, including theories of delinquency and crime, has focused on men. Given that boys and men do in fact account for the vast majority of crime, especially serious and violent crime, this is understandable. Nevertheless, female-perpetrated crime has increased substantially in recent decades, which has resulted in a corresponding proliferation of research focused on female delinquency. That said, though research on JI women in general has increased, non-White JI women have still received scant attention in the literature (Brown et al., 2014). Furthermore, until recently, Black JI women were the only historically marginalized group that had received any consideration from feminist criminologists. Encouragingly, women of Hispanic origin have begun to receive attention in the correctional literature, as have Indigenous girls and women, especially in Canada and New Zealand (see Dell, Lyons, Grantham, Kilty, & Chase, 2014; Maher, 1997).

At present, correctional scholars have yet to come up with comprehensive, testable theoretical explanations for the why non-White females engage in criminal behaviour. Existing theories can be broadly subdivided into one of three theoretical positions: (a) the gender/race similarities hypothesis, (b) the gender/race difference hypothesis, and (c) the double jeopardy effect. The first perspective argues that the

pathways leading to involvement in the criminal justice system are similar, regardless of gender, race, and ethnicity, and the second perspective is essentially the reverse. The final perspective, the double-jeopardy effect, argues that Black girls and women are at the greatest disadvantage as a result of this race/gender combination, and experience the most systemic discrimination (Chesney-Lind, 1996; Rafter, 1985; Simpson, 1989). Bloom, Owens and Covington (2002) have posited that an even more disadvantaged group – *poor* women belonging to a historically marginalized group – exists and have argued for the existence of a triple-jeopardy effect. Taken at face value, these jeopardy effects are consistent with the tenets of multiracial feminism and the intersectionality paradigm.

Currently, there are only two theories – the masculinity model and the racialized gender stereotype expectation model – that have attempted to explain how intersectionality perspectives can be applied to explain criminal behaviour. Unsurprisingly, both of the perspectives are consistent with the gender/race difference hypothesis (Chesney-Lind, 1996). Conversely, the personal, interpersonal, and community-reinforcement (PIC-R) model (Andrews & Bonta 2010) and the developmental life-course perspective (Yoshioka, & Noguchi, 2009) espouse the gender/race similarities hypothesis. Interestingly, other models (e.g., the integrated structured-life course model) includes elements from both perspectives. Readers interested in understanding how each of these models explains crime are encouraged to see Brown et al. (2014) for a more detailed review.

Intersectionality in the correctional setting. Justice involved women are receiving increasing attention in the correctional literature as are JIPs who identify as belonging to a historically marginalized group. However, research examining the impact

of race and gender has largely failed to address these questions simultaneously. With a few notable exceptions (e.g., a correctional strategy designed specifically for Indigenous JI women in Canada; see Derkzen, Harris, Wardrop, & Thompson, 2017), the majority of corrections-focused research has examined either race or gender, ignoring the multiplicative effects that the intersection of these factors engenders. Most likely, the intersectionality paradigm has important implications for non-White JI women at all junctures of the correctional process (i.e., from arrest to release). However, given this dissertation's focus, issues related to risk assessment and prediction of recidivism receive particular attention.

Assessment tools developed and validated on aggregate samples often work quite well for the majority (i.e., White, male) JIPs. However, these tools are potentially less suitable for JIPs belonging to marginalized subpopulations as they may not adequately capture the risk and need factors that are of particular relevance for these groups. The underlying question here is what impact could intersecting sources of discrimination and oppression (e.g., being female, belonging to a historically marginalized group) have on the validity of mainstream assessment tools? To date, few empirical investigations have been conducted to examine the multifactorial impact of race and gender on risk assessment. Of the existing research, some of the earliest studies in this area were completed in Canada. For example, in 2002, Blanchette, Verbrugge, and Wichmann examined the validity of the Custody Rating Scale (CRS; Luciani, Motiuk, & Nafekh, 1996) using a sample of Indigenous and non-Indigenous women. Findings indicated that the CRS tended to over-classify women of Indigenous decent (i.e., Indigenous women were underrepresented in minimum security and overrepresented in medium and

maximum security). Blanchette et al., (2002) also found that the custody level that Indigenous women were placed in according to their CRS scores did not correspond to proportions of institutional violations and that CRS scores for Indigenous women were of little value in predicting outcomes of interest (r= .03). Positively, these concerns about the suitability of the CRS for use with JI women and Indigenous inmates in particular led to the development and validation of a new classification tool for women, the Security Reclassification Scale for Women (SRSW; Blanchette & Taylor, 2007).

Holsinger, Lowenkamp, and Latessa (2006) conducted in a similar study in the United States. Holsinger and colleagues (2006) examined the predictive validity of the LSI using White and Native American JI men and women. Notably, unlike the earlier study conducted by Blanchette and colleagues in 2002, Holsinger et al., (2006) were able to fully disaggregate analyses by race and gender. Overall, results of this study were consistent with the findings of Blanchette et al., (2002); LSI scores only marginally predicted rearrests for Native American men and the relationship found between LSI scores and rearrests for Native American women, though not statistically significant, was actually in the opposite direction to what was expected. Furthermore, the relationship between LSI-dictated risk group (i.e., low, moderate, and high) and reoffence rate did not follow a linear trend for Native American women; results indicated that 50% of the low risk group had reoffended whereas only 20% of those classified as medium risk reoffended. Although these results should be considered cautiously owing to the small sample size for the Native American women group (n= 40), they nevertheless underscore the need for further research examining the applicability of risk assessment tools

developed and validated using samples comprised of primarily White men to historically marginalized subpopulations.

Chapter 3: Extant Research on the DRAOR

This chapter summarizes the research conducted to date on the DRAOR with an emphasis on findings germane to the present project. For a comprehensive description of the subscales and items comprising the DRAOR, see Chapter 4. Notably, all research on the DRAOR to date has been conducted using samples of JIPs drawn from the New Zealand Department of Corrections and the Iowa Department of Corrections, and the majority has used samples of adult general JIPs. However, several very recent studies (i.e., completed in 2015 or later) have sought to validate the DRAOR with youth JIPs (Fergusson, 2015; Muirhead, 2016), and to extend extant findings regarding the use of the DRAOR with JIPs who committed sexual offences (Averill, 2016), and violent offences (Lowenkamp, Johnson, Trevino, & Serin, 2016). Relevant evidence from these studies is described below and is summarized at the end of the section in Table 1.

Implementation and Validation

Following its design and development (Serin, 2007), initial pilot studies of the DRAOR were conducted by Tamatea and Wilson (2009), and Wilson and Tamatea (2010) in New Zealand. The DRAOR was first piloted on a small sample ($N = 59$) of New Zealand probationers (Tamatea & Wilson, 2009). Results from these pilot studies indicated that the DRAOR had adequate reliability and validity and that it was straight-forward and user-friendly to administer. The DRAOR was also piloted in the US using a sample of JIPs in Iowa State Probation and Parole; findings regarding predictive accuracy were consistent with those obtained in New Zealand, with *AUC*s of .65, .60, .65, and .67

for DRAOR Total, Stable Risk, Acute Risk, and Protective Factor scores respectively (Serin & Prell, 2012).

Hanby (2013) was the first to examine the psychometric properties and predictive abilities of the DRAOR using a large sample (*N* = 3,498). This research was conducted using New Zealand parolees and roughly half of the JIPs in the sample were Maori, which was consistent with the pattern of ethnic distribution seen across New Zealand parolees at the time. Overall, Hanby's (2013) results indicated that the DRAOR demonstrated acceptable psychometric properties, though she identified a need for future research to seek to refine the Acute subscale to improve its reliability. In addition to testing the model fit of the factor structure described by the original DRAOR subscales, Hanby (2013) also investigated two prospective models which were based on the results of analyses performed on the initial DRAOR assessment scores she collected. Despite mixed findings regarding which of the three models tested provided the best for the data (fit differed according to the timing of the assessment [e.g., initial assessment, or last]), Hanby indicated that she was not recommending changes to the DRAOR given that the original subscales still demonstrated good psychometric properties and utility in practice. For example, the Stable and Protective subscales demonstrated high internal reliability (Cronbach's α = .81 & .84), and the Acute subscale displayed moderate reliability (α = .62). Results also indicated that both the dynamic items and protective factors included in the DRAOR were predictive of any reconvictions during the two-year follow-up period, with *AUC*s of .66, .72, and .67 for the Stable, Acute, and Protective subscales respectively. Additionally, DRAOR Total scores predicted both any reconviction, and criminal reconviction, with *AUC*s of .71 and .66 respectively. Hanby (2013) also found

evidence that more proximal assessments improved prediction, as evidenced by an increase in predictive power between first (Cohen's *d* = .53) and last DRAOR assessment (*d* = .79). DRAOR Total scores demonstrated predictive accuracy for both European and Maori JIPs, and for both subsamples, predictive accuracy was improved as a function of reassessment (e.g., Cohen's *d* increased from .47 to .69 between initial and last DRAOR assessment for Maori JIPs). Finally, with respect to reoffending, Hanby found that the DRAOR predicted recidivism above and beyond the prediction provided by the RoC*RoI (Bakker, Riley, & O'Malley, 1999) which is the static risk assessment tool utilized by the New Zealand Department of Corrections. That said, the model that yielded the best overall predictive ability (*AUC* = .77) resulted from a combination of static indicators (RoC*RoI scores) and dynamic items (the DRAOR Stable and Acute scores). Collectively, Hanby's (2013) findings have a number of important implications. Namely, they suggest that the DRAOR, in its original form, demonstrates acceptable psychometric properties, and that it can accurately predict both general reconvictions (e.g., technical violations, breaches), and new criminal convictions. Next, her findings demonstrate that the DRAOR is sensitive to change, and that dynamic factors can predict incrementally over static risk variables. They also provide important preliminary evidence that more proximal assessment scores predict recidivism more accurately, when compared to baseline assessment scores. Finally, they lend credence to the previously discussed argument that risk assessment instruments based on sound theory and empirically derived risk and protective factors can predict outcome free of bias across racial groups.

Several other studies have examined various aspects of the DRAOR using samples of JIPs drawn from the New Zealand correctional population. Yesberg and

Polascheck (2015) examined a sample of high-risk male JIPs (*N* =299) who received sentences of two years or longer, and who had been assessed as having a high risk of recidivism within five years based on RoC*RoI assessment scores. The authors were interested in assessing whether first DRAOR assessments following release from prison predicted recidivism and exploring the structure and broader predictive capabilities of the DRAOR. Similar to the sample used in Hanby's (2013) analyses, this sample was predominantly Maori (60%), with the remaining thirty percent identifying as either European (30%), Pasifika (7%), or Other (3%). Noteworthy, in their examination of the structure of the DRAOR, Yesberg and Polaschek (2015) also found evidence suggesting that the three subscale (original) DRAOR structure established by Serin and colleagues did not provide adequate fit for their sample's data ($\chi^2(19) = 394.21$, $p < .011$; CFI = .84; RMSEA = .076). A subsequent principal components analysis suggested a four-subscale structure, which essentially involved further subdividing the original Acute subscale into two separate subscales, an Internal Acute subscale, and an External Acute subscale (Yesberg & Polaschek, 2015). Given that the four-subscale model (which they refer to as the *new structure*) provided a better fit for the sample data, Yesberg and Polaschek (2015) used it for their analyses, although they noted that results would have been quite similar if the original three-subscale structure had been used. To assess the DRAOR's ability to predict recidivism, the authors examined four indices of recidivism (breaches of parole, new convictions [excluding breaches], any violent conviction, and any conviction leading to imprisonment) in relation to DRAOR Total and subscale scores for both the original and new structure models. For both models, *AUC*s for the Protective subscale were significant (and equal) when predicting breaches of parole, new convictions, and

reimprisonment, (*AUC*s = .58, .60., and .62 for each of the three indices of recidivism respectively; Yesberg & Polaschek, 2015). Similarly, scores on the Stable subscale (for both models) predicted new convictions, and reimprisonment, as did DRAOR Total scores (*AUC*s between .59 and .62). However, although Acute subscale scores (from the original structure) predicted both new convictions and reimprisonment (*AUC*s = .57 and .60, respectively), scores on the Internal and External Acute subscales (from the new structure) did not predict any of the four recidivism outcomes. Although the authors were not able to further explore these non-significant findings, they did consider the possibility that it might be due to the fact that the assessments took place immediately following release and that in some cases, recidivism did not occur until six months later. Some (e.g., Hanson & Harris, 2000) have argued that acute factors may have little ability to predict risk in the longer-term, and that we should instead expect greater predictive validity when examining acute item ratings closer to the time of recidivism. Next, Yesberg and Polaschek (2015) used a series of logistic regression analyses to assess the unique contribution of each subscale in the prediction of recidivism. Results suggested that the DRAOR (both the original and the new structure) was only able to distinguish between men who were reconvicted following a new offence and those who were not; models for both original and new structures were significant for the prediction of reconviction, with $\chi^2(3, N = 299) = 9.37$, $p = .025$, and $\chi^2(4, N = 299) = 10.88$, $p = .028$, but not for any of the other three recidivism outcomes. Importantly, the model based on the new structure explained slightly more of the variance in reconviction (36-51% of the variance, versus 31-44%), although both models correctly classified 70% of the cases. Also, results revealed that for the new structure model only, the adjusted Stable subscale

made a significant unique contribution to the model. Finally, the authors examined the incremental predictive validity of the DRAOR using a series of hierarchical logistic regressions. They focused on the new stable subscale's incremental predictive power in the prediction of reconvictions, and DRAOR Total score in the prediction of reconvictions and reimprisonment. Both the new Stable subscale and DRAOR total scores demonstrated incremental predictive validity over and above that of the RoC*RoI, when added in the second block of the regression model, which further substantiates the need to integrate dynamic risk and protective factors in risk assessment practices.

In another examination of the DRAOR using New Zealand data, Yesberg, Scanlan, Hanby, Serin, and Polaschek (2015) sought to assess the predictive utility of the DRAOR with JI women and to provide evidence for the concept of gender-neutrality in risk assessment. Using a matched sample of 133 women and 133 men on parole in New Zealand, they were able to determine that DRAOR total scores predicted recidivism for both women, $X^2(1, N = 133) = 13.90$, $p < .001$, and for men $X^2(1, N = 133) = 5.15$, $p = .023$. Next, Yesberg and colleagues (2015) used Cox regression survival analysis to assess whether the three subscales in combination predicted recidivism, and to further explore the unique contributions of each DRAOR subscale. Separate models were run for men and women, and interestingly, when the three subscales were entered together in the first block of predictor variables, they significantly predicted recidivism for women, but not for men. Moreover, for women, the Acute subscale made a statistically significant unique contribution (e.g., it predicted recidivism above and beyond the contributions of the Stable and Protective subscales), with Wald(1) = 4.50, $p < .05$, $OR = 1.226$. The associated hazard ratio statistic indicates that women's likelihood of recidivism increases

by 22.6% for every 1-point increase in Acute scores (Yesberg, et al., 2015). For men, no one subscale made a significant unique contribution. Finally, the authors examined the incremental predictive validity of DRAOR Total scores for both genders and the Acute subscale scores (for women only) in relation to static risk, as assessed by the RoC*RoI, in the prediction of recidivism. The addition of DRAOR Total scores improved prediction above and beyond that of the RoC*RoI for women (Wald(1) =8.33, $p < .01$, $OR = 1.081$), but not for men, indicating that for women, for every 1-point increase in DRAOR Total score, likelihood of recidivism increased by 8.1%. Similarly, for women, the Acute subscale made its own unique contribution to the prediction of recidivism above and beyond the RoC*RoI, with $OR = 1.234$. Taken collectively, findings Yesberg and colleagues' (2015) findings suggest that the risk assessment tools like the DRAOR developed on JI men do generalize to JI women. Interestingly, the authors actually found that the DRAOR predicted recidivism more robustly for JI women than men. Noteworthy, the findings that the Acute subscale (which contains items pertaining to substance use, relationship problems, etc.,) was particularly salient for the women in the sample is consistent with extant literature (e.g., Van Voorhis et al., 2010). In sum, Yesberg and colleagues' (2015) findings regarding the predictive validity of the DRAOR with JI women lends support to the gender-neutrality of the tool. However, there remains the possibility that the DRAOR's predictive utility with women could be further improved through the inclusion of additional gender-specific variables.

In a recent study, Lloyd (2015) proposed and tested a framework for examining the proximity hypothesis (i.e., that scores on more proximal assessments will predict outcome more accurately than scores from more distal assessments). Lloyd (2015) used

longitudinal multiple-reassessment reentry data and argued that Cox regression survival analysis with time-varying predictors is an ideal statistical model when the goal is to compare the relative value of baseline, and later assessment scores. Using a large sample of paroled New Zealand JIPs ($N > 3{,}000$) who were regularly assessed on the DRAOR by their supervision officers (represented by $N > 97{,}000$ assessments), Lloyd's findings suggested that it is important for correctional staff to reassess JIPs using dynamic instruments such as the DRAOR, as doing so yielded substantive benefits in terms of recidivism prediction. Specifically, his findings provided strong support for the proximity hypothesis, demonstrating that reassessment scores provided incremental predictive validity above and beyond baseline scores, and moreover, that more proximal assessments predicted incrementally over averages of earlier scores.

As alluded to above, the DRAOR has also been validated using U.S. samples. Chadwick (2014) sought to validate the DRAOR using a sample of Iowa probationers and parolees (N= 391). Largely concurrent with the above studies, Chadwick's (2014) findings suggested that overall, the DRAOR possessed adequate psychometric properties (e.g., factor structure, inter-item correlations, item distributions, and internal consistency), with the exception of the Acute subscale, which demonstrated poor internal consistency. Chadwick (2014) found that amalgamating the original Stable and Acute subscales into a single Risk scale led to improvements in internal consistency, and ultimately a more concise model than the original DRAOR model, thus, both this, and the original three subscale DRAOR were examined in subsequent predictive validity analyses. With respect to predictive validity, Chadwick (2014) hypothesized that DRAOR Total and subscale scores would be able to differentiate between recidivists and

non-recidivists and found partial support for this hypothesis. Chadwick (2014) examined the predictive accuracy of the DRAOR using ROC analyses and found that the DRAOR and its subscales consistently demonstrated the ability to predict technical violations and any recidivism. The Stable domain and Total score produced the largest effect (*AUC*s between .61 – .62). Contrary to expectations, the DRAOR did not predict whether an individual JIP was likely to reoffend (e.g., new criminal convictions). These findings are inconsistent with previous research (e.g. Hanby, 2013, Yesberg et al., 2015), which found support for the prediction of criminal reconviction, with *AUC*s ranging from .66 to .72 (Hanby, 2013). However, these findings need to be understood in view of a substantial caveat – Chadwick (2014) noted that not all supervision officers received the same training, and thus, proficiency and comfort in administering the DRAOR could not be ensured. To address this limitation, he conducted additional prediction analyses, and significant differences in predictive accuracy based on training status were found. Importantly, Chadwick (2014) found that DRAOR scores as assessed by formally trained officers were moderately related to recidivism (significant *AUC*s ranging from .62 to .69), whereas DRAOR scores from informally trained officers did not evidence a significant relationship with recidivism (non-significant *AUC*s ranging from .52 to .58).

In another American study, Smeth (2013) examined the predictive validity of dynamic risk and protective factors, as assessed by the DRAOR, using a sample of 212 adult male sex offenders. Hers was the first study that specifically sought to validate the DRAOR as a sex offender assessment tool, and therefore represents a unique contribution to the DRAOR literature. Results indicated that DRAOR Total scores and all subscale scores predicted rule violations (*AUC*s ranging from .61 to .69), although they did not

predict sexual recidivism (non-significant *AUC*s ranging from, .46 to .53; Smeth, 2013). Next, Smeth (2013) used a series of Cox regression survival analyses to assess whether DRAOR total and subscale scores were related to survival time for sexual recidivism and parole failures. Again, results indicated that DRAOR Total and subscale scores did not significantly predict sexual recidivism, although they were predictive of time to parole violation. Finally, Smeth (2013) examined the incremental predictive validity of dynamic risk and protective factors (assessed using the DRAOR), above and beyond Static-99R scores using a series of hierarchical cox regression analyses. Her findings indicated that the addition of DRAOR scores incrementally enhanced predictive validity above and beyond the predictive accuracy achieved based on static factors alone. Although Smeth's results did not offer strong support for the predictive utility of the DRAOR for sexual recidivism, it should be noted that low prevalence rates for sexual recidivism may have negatively impacted the likelihood of finding significant results. Nonetheless, these results offer further support for the overall ability of dynamic risk and protective factors to predict general recidivism, and to predict incrementally over and above static risk variables.

Similarly, Averill (2016) sought to investigate the predictive validity of the DRAOR using a cohort of sexual offenders (N = 851) on community supervision in New Zealand. Averill (2016) examined the DRAOR's ability to predict sexual, violent, and general recidivism, as well as incremental predictive ability above and beyond the RoC*RoI, and findings were largely consistent with those of Smeth (2013). Like Smeth (2013), Averill found that the DRAOR had good predictive validity with respect to time to violent, general, and administrative offences for the sexual offenders in the sample and

yielded incremental predictive validity over measures of static risk (the RoC*RoI). Likewise, DRAOR scores were not significantly predictive of time to sexual recidivism, after controlling for static risk.

Both Fergusson (2015) and Muirhead (2016) investigated the predictive utility of the DRAOR with younger JIPs using samples of youths (ages 17-19) serving community supervision orders in New Zealand. Both studies provided additional evidence for the necessity of updating reassessing JIPs using the DRAOR. Fergusson (2015) found that although initial (baseline) scores offered inconsistent results with respect to the prediction of recidivism, proximal risk scores predicted recidivism with moderate to high accuracy. Likewise, Muirhead (2016) found that more proximal DRAOR assessments predicted recidivism more accurately. Importantly, youth JIPs represent yet another correctional subgroup; thus, both studies provide additional evidence for the applicability of the DRAOR across JIP subpopulations.

Finally, Lowenkamp, Johnson, Trevino, and Serin (2016) examined if key acute risk factors (i.e., those included in the Acute subscale of the DRAOR) could be used to forecast violent rearrests in a sample of American federal probationers. Data for the study included federal supervision records assembled from the Probation and Pretrial Services Office's internal case management system (Probation and Pretrial Services Automated Case Tracking System) in addition to other relevant sources. The final sample was comprised of 4,116 JIPs who had been arrested for a serious violent offence (i.e., homicide, attempted homicide, sexual assault, robbery, and felonious assault). In addition to Acute DRAOR scores, Lowenkamp and colleagues (2016) developed a static violence classifier that represented a JIP's risk of committing a violent offence. Individuals in the

sample were considered a higher risk for violence if they had a PCRA (Post Conviction Risk Assessment; Johnson, et al., 2011) above a pre-determined threshold. Also, the authors created a dichotomous variable representing age of onset of offending behaviour, with JIPs whose first arrest occurred prior to age 18 being assigned a value of 1, and those whose first offence occurred in adulthood receiving a score of 0. Cox regression analyses indicated that once dynamic acute risk factors were entered into the model, the effect of the violence classifier became non-significant though the age at onset variable retained significance as a predictor of time to failure. This suggests that the acute dynamic risk factors were able to capture risk for violence. Moreover, DRAOR Acute scores of four or greater significantly predicted failure, and hazard ratios (*OR*s) showed a general upward trend, suggesting that increases in DRAOR Acute scores were mirrored by increased likelihood of failure. Finally, Lowenkamp and colleagues (2016) assessed the predictive utility of each of the DRAOR Acute items independently to evaluate if any of the risk factors were especially predictive of failure (i.e., arrest for a violent offence). Regression analyses indicated that three Acute items (anger/hostility, access to victims, and negative mood), were significantly related to time to failure. For each of these three factors, having a *slight/possible problem* (a score of 1) versus having *no problem* (a score of 0) resulted in a significant decrease in survival rates, with *OR* values of 1.90, 1.60., and 1.41 for anger/hostility, access to victims, and negative mood, respectively. Moreover, for the anger/hostility and access to victims items, a shift in ratings from having a *slight/possible* problem (a score of 1), to having a *definite problem* (a score of 2) resulted in a further significant increase in failure rates. For anger/hostility, the shift from a rating of 1 to a 2 produced a hazard ratio of 3.08 ($p < .001$*),* and for access to victims, a

shift from a rating of 1 to a 2 corresponded with a hazard ratio of 3.04 ($p < .001$). Lowenkamp et al.'s (2016) findings are informative, as they demonstrate that overall increases in Acute risk are predictive of violent rearrest, and that scores on the anger/hostility, access to victims, and negative mood variables are especially predictive of failure rates. Their non-significant findings regarding the independent predictive abilities of the other Acute items are likewise informative; factors like substance use and interpersonal problems have been shown to be predictors of recidivism in other studies (e.g., many gender-related studies like Van Voorhis, et al., 2008, 2010), therefore it is noteworthy that they did not inform the likelihood of violent rearrest here. More generally, this study provides further support for the argument that dynamic risk factors, and acute factors in particular, need to be consistently addressed with clients. Given how quickly acute factors are liable to change, frequent reassessments are imperative to improve client outcomes with higher risk individuals.

Table 1

Associations Between DRAOR Scores and Recidivism Outcomes Reported in Previous DRAOR Research

Source	Sample	Outcome	Stable		Acute		Protective	
			r	*AUC* [95% CI]	*r*	*AUC* [95% CI]	*r*	*AUC* [95% CI]
Serin & Prell (2012)	General offenders [I]	Technical violations or new offence	.23	.60 [.56, .64]	.25	.65 [.61, .69]	-.28	.67 [.63, .70]
		New offences only	.12	-	.11	-	-.15	-
Hanby (2013)	General offenders [NZ]	Technical violations or new offence	.29	.66 [.65, .68]	.38	.72 [.70, .74]	-.29	.67 [.65, .68]
		New offences only	.21	.62 [.60, .64]	.29	.67 [.65, .69]	-.21	.62 [.60, .64]
	Maori offenders	Technical violations or new offence	.26	.65 [.62, .67]	.35	.70 [.68, .72]	-.26	.65 [.62, .67]
Smeth (2013)	Male sexual offenders [I]	Technical violations	.33	.69 [.62, .78]	.27	.65 [.57, .74]	-.18	.61 [.52, .69]
		Sexual Recidivism	.01	.51 [.26, .75]	.05	.46 [.30, .75]	.01	.46 [.21, .71]
Chadwick (2014)	General offenders [I]	Technical violations or new offence	-	.62 [.56, .67]	-	.59 [.54, .65]	-	.58 [.52, .64]
		New offences only	-	.52 [.42, .62]	-	.53[.43, .63]	-	.56 [.47, .66]
Yesberg & Polaschek (2015)	High-risk male offenders [NZ]	Technical violations or new offence	-	.59 [-,-]	-	.60 [-,-]	-	.62 [-,-]
		New offences only	-	.61 [-,-]	-	.57 [-,-]	-	.60 [-,-]
Fergusson (2015)	Youth offenders [NZ]	Criminal reconviction		.61 [.49, .73]^{ns}		.60 [.48, .71]^{ns}		-.60[.48,.71]^{ns}
Averill (2016)	Sexual offenders [NZ]	Technical violations or new offence	.33	-	.37	-	-.33	-
		Sexual recidivism	.02 ^{ns}	-	.07	-	-.06^{ns}	-
Muirhead (2016)	Youth offenders [NZ]	Any reconviction	-	.70 [.64, .76]	-	.69 [.63, .74]	-	-.68 [.62, .74]

Note. DRAOR = Dynamic Risk Assessment for Offender Re-entry (Serin, 2007). Stable = DRAOR *Stable* subscale. Acute = DRAOR *Acute* subscale. Protective = DRAOR *Protective* subscale. *r* = point-biserial correlation. *AUC* = area under the curve statistic. CI = confidence interval. [I] = Iowa sample. [NZ] = New Zealand Sample. "-" = results not reported. *ns* = not significant. Four DRAOR studies (Lloyd, 2015; Lowenkamp et al., 2016; Yesberg et al., 2015; Tamatea & Wilson, 2009) are omitted as no compatible *r* and *AUC* values were reported.
$p < .05$ unless otherwise indicated (*ns*).

Summary and Conclusions

Considered collectively, existing DRAOR studies are encouraging, yet suggestive of a need for continued refinement and validation. For example, several of the above studies (e.g., Chadwick, 2014; Hanby 2013; Yesberg & Polaschek, 2015) found that the current/original three-factor structure of the DROAR did not provide an ideal fit for their data. Although none of these authors definitively concluded that the DRAOR should be restructured, further factor analytic research is clearly merited. Structural uncertainties aside, the DRAOR has demonstrated that it can be reliably used in prediction. Results generally indicated that the DRAOR accurately predicts a variety of recidivism outcomes (e.g., general reconvictions, breaches of parole, violent rearrests), and moreover, that it can incrementally improve prediction above and beyond estimates based on static risk alone. Likewise, there is preliminary evidence to suggest that the DRAOR is predictive for JIPs of various races, as well as for both male, and female JIPs, and can potentially be considered race- and gender-neutral. Notably, extant research on the DRAOR has not offered clear support for its conceptualization of protective factors. For instance, Chadwick (2014), and Smeth (2013) found no evidence of an interaction effect between protective factors and JIPs' level of risk, which is what would have been expected if the DRAOR items were having a genuinely protective effect as defined by Jones et al. (2015). Nevertheless, the majority of findings regarding the DRAOR are encouraging regarding the promotive effect of protective factors and the incremental predictive validity of the DRAOR above and beyond static risk, as findings from several studies (e.g., Chadwick, 2014; Hanby, 2013; Lloyd, 2015) have demonstrated that it significantly improves the prediction of recidivism. Finally, findings from several of the above

DRAOR studies such as Lloyd (2015) provide support for the proximity hypothesis; that is, they substantiate the importance of consistently reassessing JIPs' dynamic (e.g., changeable) risk and protective factors, as more recent assessments scores contributed to increased accuracy in the prediction of recidivism.

Review of Purpose

The purpose of the current project is to advance the field's understanding of the predictive utility of dynamic risk and protective factors with JI women from diverse racial and ethnic backgrounds. Concurrently, this project aims to extend the validation literature on the DRAOR by (a) clarifying its factor structure and assessing measurement invariance for across racial groups and over time (Study 1), (b) examining base rates and survival trajectories for each subgroup of JI women in relation to different outcomes of interest (e.g., any return, technical violations, and new offences; Study 2), and (c) exploring any nuances in predictive validity at the domain and item level with particular attention being accorded to potential promotive/protective factors and how findings might pertain to intersectionality theory (Study 3).

Chapter 4: Common Methodology

This chapter describes the overall sample, procedure, measure(s) and outcome variables that are relevant for the three studies presented in Chapters 5, 6, and 7. This information was presented in a single chapter to reduce redundancy later in the document. Study-specific components of methodology (i.e., purpose, hypotheses, variations in sample, and proposed analyses) will be presented at the beginning of each study as appropriate. Note that this research received ethics clearance from the Carleton University Research Ethics Board-B (Protocol # 106316 14-073; see Appendix A).

Sample

At this juncture, it is important to acknowledge several shortcomings with respect to the terminology used to describe the gender and ethnicity of the individuals included in the sample. The dataset provided by the IDOC contained only a binary male/female classifier, which reflects biological sex rather than gender and does not take other gender identities into account.[4] Thus, although the term gender is used in this dissertation, it should be noted that it is used in this limited capacity. How individuals' race and ethnicity were coded in the dataset also presented challenges. Individuals were classified both by race (American Indian/Alaska Native, Asian or Pacific Islander, Black, or White), and by ethnicity (Hispanic or Non-Hispanic).[5] As the ethnicity variable only distinguished between individuals identifying as Hispanic or non-Hispanic, it was necessary to blend race and ethnicity variables to create different study groups (described below). As the final study groups were more reflective of the IDOC's categorization of individuals' race rather than self-identified ethnicity, the term race was utilized in this dissertation.

The initial sample for this project comprised 3,352 justice involved women on community supervision orders in the state of Iowa.[6] Information about race and ethnicity

---

[4] The Canadian Institutes of Health Research (CIHR, 2019) defines gender as the socially constructed roles, behaviours, expressions and identities of girls, women, boys, men, and gender diverse persons. Gender influences and individual's perception of themselves and of others, and influences not only interactions among people, but also the distribution of power and resources in society.

[5] As per Statistics Canada (2017), the term 'race' is typically used to refer to genetically imparted physiognomical features, including skin colour, and is no longer considered suitable terminology. Ethnicity, the preferred term, is multidimensional, and includes aspects such as race, origin or ancestry, identity, language and religion, and may also include more subtle dimensions such as belief systems and diverse elements of culture.

[6] The data set received from Iowa also contained information for more than 15,000 men. Descriptive information is not provided here, as with the exception of a small subset of men employed as a comparison sample in Study 3, male offenders are not considered in this dissertation.

was not provide for 31 women, and these cases were removed from the sample. Of the remaining 3,321 women, nearly two-thirds (64.9%, n = 2,155) had at least two DRAOR assessments, with 3.7 (4) DRAOR assessments completed per woman on average. Cases were then scanned to ensure that all women in the sample had at least one valid DRAOR assessment on file.[7] Frequency analyses identified 143 cases without a valid assessment score; these 143 cases were removed, resulting in a final sample of 3,178 women. All JIPs in the final sample also had at least one assessment score on the Iowa Violence and Victimization Instrument (IVVI), Iowa's static risk assessment tool.

Demographic information for the full sample is shown in Table 2 below. A small portion of the sample (n = 113, 3.6%) identified as Hispanic, and the remaining 96.4% identified as non-Hispanic (n = 3,065). Non-Hispanic women were further categorized by race. The vast majority of the non-Hispanic sample was White (n = 2,418, 76.1%), with Blacks representing the next largest group (n = 560, 17.6%), and American Indian/Alaska Natives and Asian or Pacific Islanders (n = 75 and n = 15 respectively), accounting for less than 3% collectively. For the purposes of this project, this group will henceforth be referred to as Other. White women were oldest overall (M = 35.1, SD = 9.8), followed by women in the Other category (M = 34.3, SD = 9.0), and Black women (M = 32.9, SD = 10.5). Hispanic women were the youngest (M = 31.6, SD = 9.4). White women were significantly older than Black and Hispanic women ($t(798.85) = 4.53$, $p < .001$ and $t(123.56) = 3.81$, $p < .001$ respectively) but not Other women. Across all races, women in the sample were most likely to be single and to have graduated from high school.

[7] Multiple DRAOR assessments were required for longitudinal measurement invariance analyses in Study 1. One assessment (the baseline DRAOR score) was utilized in Studies 2 and 3.

Table 2

Age, Marital Status, and Education Level of Iowa JI Women

	All Women (*N* = 3,178)	White (*N* =2,418)	Black (*N* =560)	Hispanic (*N* =113)	Other (*N* =87)
Age					
18-20	98	60	28	†	†
21-30	1170	833	260	53	24
31-40	1091	880	141	33	37
41-50	640	437	80	18	15
51+	269	208	51	†	6
Marital Status					
Single	1700	1179	411	67	43
Common-law	27	22	†	†	†
Married	456	380	46	15	15
Divorced	461	411	26	13	11
Separated	143	124	11	†	†
Widowed	44	36	7	†	†
Highest Education					
Grade 8 or lower	77	64	8	†	†
Some high school	512	333	139	24	15
High school diploma/GED	2066	1613	326	70	57
Post-secondary or other vocational training	311	260	33	10	8
Unknown	212	149	54	†	6

Note. The Other group combines American Indian/Alaska Natives and Asian or Pacific Islanders. † = cells with *N* of 5 or less are supressed.

Information regarding static and dynamic risk levels for the women in the sample is provided in Table 3. In Iowa, an JIP's level of supervision (frequency of contact) is determined in accordance with their level of static risk, as assessed by the IVVI. Administrative supervision represents the lowest intensity supervision, followed by minimum, low normal, and high normal. Intensive supervision is reserved for JIPs who pose the most serious recidivism risk. Accordingly, the sample used in this dissertation can be described as high risk, with more than two thirds (68.3%) of the sample supervised at the high normal or intensive level. Nearly half (49.4%) of Other women and

44.5% of Black women were supervised at the intensive level. Dynamic risk was assessed using the DRAOR. DRAOR Total scores were divided into four risk bins, based on similar frequency of recidivism (see Perley-Robinson, Chadwick, & Serin, 2020). Most women (74.5%) were rated as having moderate or moderate-high levels of dynamic risk. The largest proportion of White and Other women were assessed as moderate-high risk, whereas Black and Hispanic women were most likely to be rated as moderate dynamic risk. Although the overall sample can be considered high risk on account of the large proportions rated as having above average levels of static and dynamic risk factors, very few women (1.1%) fell into the highest risk category on the DRAOR.

Table 3

Level of Supervision and Dynamic Risk Ratings as Assessed by the Iowa Violence and Victimization Instrument (IVVI) and the Dynamic Risk Assessment for Offender Re-Entry

	All Women (*N* = 3,178) *n* (%)	White (*N* =2,418) *n* (%)	Black (*N* =560) *n* (%)	Hispanic (*N* =113) *n* (%)	Other (*N* =87) *n* (%)
Level of Supervision					
Administrative	31 (1.0)	24 (1.0)	5 (0.9)	†	†
Minimum	56 (1.8)	52 (2.2)	†	†	†
Low Normal	922 (29.0)	721 (29.8)	146 (26.1)	31 (27.4)	24 (27.6)
High Normal	918 (28.9)	703 (29.1)	157 (28.0)	39 (34.5)	19 (21.8)
Intensive	1251 (39.4)	918 (38.0)	249 (44.5)	41 (36.3)	43 (49.4)
Dynamic Risk Rating					
Low–Moderate	754 (24.4)	606 (25.1)	120 (21.4)	28 (24.8)	20 (23.0)
Moderate	1132 (36.6)	845 (34.9)	240 (42.9)	47 (41.6)	31 (35.6)
Moderate–High	1171 (37.9)	942 (39.0)	192 (34.3)	37 (32.7)	34 (39.1)
High	34 (1.1)	25 (1.0)	8 (1.4)	†	†

Note. The Other group combines American Indian/Alaska Natives and Asian or Pacific Islanders.
† = cells with *n* of 5 or less are supressed.

Importantly, following some consideration, the Other group was removed from analyses. Given that this dissertation is concerned with examining the experiences of women belonging to distinct, yet relatively cohesive racial groups, it is important that the women in each group can be presumed to share a common background and be exposed to

similar treatment by correctional staff. It cannot be assumed that women identifying as American Indian, Alaska Native, Asian, and Pacific Islanders share similar characteristics or experience the criminal justice system in the same way, thus, they should not be lumped together for the purpose of analysis. To be included, more women belonging to each of these racial/ethnic groups would have been required. Demographic characteristics for the Other group were described above to provide a more accurate picture of Iowa's female correctional population. Following the removal of the 87 women in the Other group, the final sample consisted of 3,091 White, Black, and Hispanic women.

Procedure

Data for the present project were obtained courtesy of the Iowa Department of Corrections (IDOC) as part of a collaborative research agreement. Data were retrieved from Iowa's JIP data management system (the Iowa Correctional Offender Network; ICON), and the present dataset includes information about JIPs whose first DRAOR assessment (and any subsequent DRAOR assessments) took place between September 2014, and April 2019. DRAOR assessments were completed during individuals' regularly scheduled meetings with their parole officer, and scores from community supervision offices state-wide were saved into a larger common administrative dataset.

Measures

Sample demographic information and recidivism. The dataset received from the IDOC for use in this dissertation contained all of the basic demographic information routinely collected from JIPs when they first make contact with the correctional system. This includes each individual's age, gender, race, ethnicity, highest level of education, and marital status. The gender, and race/ethnicity variables were of greatest interest for

the present project. The data file also contained information describing index offences (e.g., class of crime and type of offence), supervision status and level of intensity of supervision (e.g., probation, parole, work release), and any subsequent changes in status. Community outcome (recidivism) data was also included in the dataset. The date, class, and type of recidivism are provided, along with any associated new charges (if applicable) for the period between first DRAOR assessment and April 2019. Violations of conditions of supervision were also recorded. For the present project, three types of recidivism, technical violations, new offences, and any return to custody, were examined.

Dynamic Risk Assessment for Offender Re-Entry (Appendix B1). The DRAOR (Serin, 2007) is a structured professional judgment (SPJ) case management instrument developed to assist community supervision officers in monitoring and managing the unique risks, needs, and strengths of their clients.[8] The DRAOR is comprised of a total of 19 dynamic items, which are further subdivided into three separate domains, or subscales. The Stable domain (six items) assesses JIPs' characteristics reflective of general criminal orientation and impulsivity that are associated with risk and that are liable to change in the slightly longer term (i.e., months to years). Conversely, the Acute domain (seven items) assesses risk factors that can change more rapidly (i.e., can change substantially within a month), such as highly transient aspects of daily life and factors linked to disinhibition. The final domain, the Protective domain (six items), assesses characteristics linked to positive valuations and social support that may buffer risk for recidivism.

[8] It should be noted that while the DRAOR was designed as a case management/ SPJ tool and is implemented in this capacity in various jurisdictions, the current research examines its utility as an actuarial assessment instrument tool.

All DRAOR items are scored similarly using a three-point scoring format. For Stable and Acute risk domain items, possible scores include 0 (*not a problem*), 1 (*slight/possible problem*), or 2 (*definite problem*), and for items in the Protective subscale, possible scores include 0 (*not an asset*), 1 (*slight/probably asset*), or 2 (*definite asset*). Given that there are six items in the Stable subscale, seven in the Acute subscale, and six in the Protective subscale, total domain scores range from 0-12, 0-14, and 0-12, respectively. DRAOR Total scores are calculated by aggregating across Stable and Acute domains, and then subtracting scores on the Protective domain. As such, DRAOR Total scores can range from -12 to +26, with higher scores indicating a higher degree of risk, and fewer protective factors. Notably, as the DRAOR is an SPJ instrument, individual items can be omitted if evidence is unreliable. Both continuous DRAOR Total scores and categorical risk bins (see Table 3 above) were used in this dissertation.

As central aims of this doctoral research are to replicate the factor structure of the DRAOR and further explore predictive validity, all of the DRAOR subdomains are of interest. That said, specific domains, and individual items within domains, are of greater relevance for certain analyses. The Acute domain contains the following seven items: substance abuse, anger/hostility, opportunity/access to victims, negative mood, employment, interpersonal relationships, and living situation. Given the rapidity with which these items are liable to change, the Acute subscale may be of particular import in identifying when recidivism is imminent. They may also play an important role in the prediction of recidivism using most proximal DRAOR assessment scores. However, given this mutability, Acute items may be less helpful when attempting to assess enduring change.

Unlike the Acute items, the Stable and Protective items are theorized to change more slowly, and correspondingly, to represent more durable change. The six factors comprising the Stable domain (peer associations, attitudes towards authority, impulse control, problem-solving, sense of entitlement, and attachment with others) are largely derived from the work of Andrews and Bonta (1995, 2010), and Hanson and Harris (2000). Stable risk factors are expected to change more gradually over time (Hanson & Harris, 2000), and may change in response to intervention, or as a function of the passage of time. As discussed earlier, the six items comprising the Protective subscale (responsiveness to advice, prosocial identity, positive expectations for reintegration success, perceiving prosocial behaviour as more rewarding than criminal behaviour, presence of prosocial supports, and willingness to follow advice from prosocial others) are conceptualized as being independent from risk (Polaschek, 2016; Serin, 2007). The development of this domain is grounded in desistance literature, and most notably, the work of Maruna (2001) and Sampson and Laub (2005). Collectively, the DRAOR items are reflected in the Transition Model of Offender Change (see Lloyd, & Serin, 2012; Serin, Lloyd, & Hanby, 2013), which describes how interpersonal factors interact with risk factors and desistance factors to moderate the initiation and escalation of criminal behaviour as well as the later desistance process. Importantly, the DRAOR represents a careful synthesis of extant best practices from both the risk assessment and desistance literature, and as such, investigating the predictive capabilities of both DRAOR total scores, and each subscale individually have potential implications for the field.

As described in Chapter 3, initial validation studies on the DRAOR using community corrections samples in both New Zealand (i.e., Hanby, 2013; Yesberg &

Polaschek, 2015; Yesberg et al., 2015) and in Iowa (i.e., Chadwick, 2014; Serin & Prell, 2012; Smeth, 2013) have demonstrated that it is a promising instrument for both initial risk assessment, and dynamic reassessment. Findings from the above studies have yielded inter-correlations among DRAOR subscales ranging from r = .61 to .74 for the Stable and Acute subscales, r = -.50 to -.71 for the Stable and Protective subscales, and r = -.47 to -.66 for the Acute and Protective subscales. Both Chadwick (2014) and Hanby (2013) assessed internal consistency for each subscale, and their findings were remarkably consistent across samples, with Cronbach's α = .81 (Stable), .62 (Acute), and .86 (Protective) and α = .81 (Stable), .62 (Acute), and .84 (Protective) for Chadwick, and Hanby respectively.

However, findings regarding the appropriateness of the intended three-factor structure of the DRAOR have been somewhat mixed (e.g., Chadwick, 2014; Hanby, 2013; Yesberg & Polaschek, 2015). That said, some (e.g., Baird, 2009) have argued that insofar as risk assessment tools are concerned, that the focus should be on maximizing prediction rather than being overly concerned with whether or not a given measure (and its subscales) assess conceptually distinct and cohesive constructs. Relatedly, Lloyd (2015) argues that the application of factor analysis to DRAOR data is necessarily a limited approach given that each of the DRAOR items was chosen for its value as an individual indicator of a personal or external variable, and subsequent categorization into subscales was determined primarily as a function of similarity of anticipated temporal relationships to recidivism. In other words, the DRAOR does not purport to capture three distinct constructs.

Nevertheless, several studies (Hanby, 2013; Yesberg & Polaschek, 2015) have provided support for the measurement invariance of DRAOR assessments across time. In both cases, the authors conducted CFAs on both baseline, and subsequent reassessment scores, and concluded that there were no significant differences in terms of model fit to their data. Briefly, measurement invariance across time signifies that there is empirical evidence demonstrating that DRAOR items measure the same underlying construct(s) at different time points (e.g., at both baseline, and subsequent assessment occasions). However, time is not the only factor germane to considerations of measurement invariance. Issues such as changes in a JIP's circumstances, variability in evaluators' understanding of item content, and differences in behavioural expectations can lead to such discrepancies across assessments (Brown et. al., 2009). Study 1 will focus on investigating both measurement invariance across time and measurement invariance across subgroups.

Second, the aforementioned studies have demonstrated that, when assessed at a single point in time (e.g., at baseline), DRAOR subscale scores are valid predictors of both technical violations and new offences while under community supervision. Key findings from initial validation research (refer to Table 1) indicated that overall, the DRAOR domain scores significantly predicted technical violations and new offences for general JIPs. Investigations of the incremental predictive validity of DRAOR subscale scores above and beyond static risk scores have also been largely successful (e.g., Chadwick, 2014; Hanby, 2013; Lloyd, 2015; Smeth, 2013; Yesberg & Polaschek, 2015). For instance, both Hanby (2013) and Yesberg and Polaschek (2014) found that DRAOR total scores added incrementally to prediction after accounting for RoC*RoI scores, and

Smeth (2013) was likewise able to demonstrate that each DRAOR subscale predicted parole violations above scores on the Static-99R in her sample of sex offenders. Somewhat perplexingly, in their study Yesberg and colleagues (2015) found that only the Acute subscale predicted recidivism among a sample of female JIPs in New Zealand, though the Acute subscale also demonstrated incremental prediction above static scores. In sum, while findings from previous research generally suggest that the DRAOR subscales are moderate to strong predictors of general recidivism, there are some inconsistencies across JIP subgroups. These inconsistencies are explored in Study 2.

Third, the available research on the DRAOR suggests that it can be used to detect and monitor change. For example, Hanby (2013) examined individual-level change in DRAOR subscale scores using multi-level modeling. Her findings demonstrated that the DRAOR was able to detect change within individuals, and that the changes occurred in the anticipated directions (e.g., JIPs who did not recidivate evidenced average increases in Protective scores and average decreases in risk (Acute and Stable) scores. Lloyd (2015) provided further evidence of the DRAOR's sensitivity to change. Using the same initial sample as Hanby (2013; a large sample of New Zealand parolees [N = 3,694]) Lloyd demonstrated not only that the DRAOR could be used to detect change at routine reassessments, but also that later assessment scores predicted incrementally over earlier scores.

Notably, the DRAOR's ability to detect and monitor change is not examined in this dissertation. Instead, the predictive ability of individual DRAOR items, a heretofore largely unexplored aspect of the tool's utility, is examined in Study 3.

Iowa Violence and Victimization Instrument (Appendix B2). The Iowa Violence and Victimization Instrument (IVVI) is a risk scale developed by the IDOC to assist in the determination of level (intensity) of community supervision and subsequent offender management. Adopted statewide in 2014 in Iowa, the IVVI is comprised of nine predominantly static items reflecting criminogenic variables associated with both violent and non-violent recidivism, such as prior offences, criminal gang membership, and current age (Prell, Vitacco, & Zavodny, 2016). The scale is typically administered at the onset of the community supervision period, and each of the nine items is considered in relation to two separate scoring scales, receiving both a Victimization score, and a Violence score. The Violence score is operationally defined as the likelihood (i.e., risk) of a conviction for any violent crime within 30 months of administration. The Victimization score is related, yet is more comprehensive, and is operationalized as the as the likelihood of a conviction for any violent offence, property offence, or misdemeanor/felony within 30 months of initial release to community supervision (Prell et al., 2016). To clarify, the Victimization score is intended to capture crimes with a quantifiable economic impact (e.g., burglary, identity theft, property crimes, etc.) felt personally by the victims. The items are straightforward, and scoring procedures are relatively simple, especially if the evaluator has access to the JIP's complete criminal record. Possible scores range from -1 to 3, although scores are not necessarily comparable across the Violence and Victimization categories (e.g., having a current age between 25 and 29 is attributed a Violence score of 2 but a Victimization score of 0; see Appendix B for complete a complete list of the IVVI items and scoring). Violence and Victimization score categories vary slightly, with Violence scores divided into low (-1 to

2), moderate (3 to 5), high (6 to 9), and very high (10+) categories, and Victimization scores divided into low (-1 to 1), low/moderate (2 to 3), moderate/high (4 to 7) and high (8+). Notably, assessment using the IVVI is a two-step process, with Victimization scores calculated first, and subsequently adjusted upwards following consideration of the Violence score (Prell et al., 2016). As noted above, IVVI scores are used to determine appropriate levels of supervision for each JIP. The five levels currently in use by the IDOC are administrative (the lowest), minimum, low normal, high normal, and intensive.

Although validation research on the IVVI is rather limited to date, results are promising. Prell, Vitacco, and Zavodny (2016) examined the basic psychometric properties and predictive power of the IVVI in a large sample (N = 1,961) of males who were on probation or released from prison to parole supervision in the state of Iowa, and concluded that the instrument was able to both accurately determine the appropriate level of supervision (intensity) and predict general recidivism. Over the 30-month follow-up period, both the Violence and Victimization scales predicted general recidivism with fair to good predictive power (*AUC*s = .65 and .63 for the Victimization scale score and Violence scale score respectively; Prell et al., 2016). Both scales were also found to be good predictors of victimization offences (*AUC* = .70 for the Victimization scale score and *AUC* = .69 for the Violence scale score), and violent offences (*AUC* = .70 for the Victimization scale score and *AUC* = .71 for the Violence scale score). As noted by Prell and colleagues (2016), the abilities of each scale to predict the corresponding type of offence (*AUC*s = .70 and .71) puts the IVVI in the same category as other more established measures like the VRAG-R and the HCR-20 with respect to predictive power. Notably, neither scale was particularly effective in the prediction of drug offences (non-

significant *AUC*s = .54 and .56 for Victimization and Violence respectively). A separate study using a sample of female probationers and parolees (*N* =987) found similar results, indicating that the IVVI is appropriate for use with women, but cross-validation studies exploring the convergent validity and generalizability of the IVVI are required (Britt, Patton, Remaker, Prell, & Vitacco, 2019).

Outcome Variables

The outcome variables of interest in this dissertation were: 1) technical violations; 2) new offences, and 3) any return. Technical violations are incurred when a JIP fails to abide by the conditions of his or her parole or probation. These misbehaviours are not, by themselves, criminal offences and may not result in reimprisonment if the violation is minor. Common examples include failing to report for a scheduled appointment, missing curfew, failing to maintain employment or attend school, and contacting a co-accused. More serious technical violations (e.g., escape, repeated failures to report) or a reoccurring pattern of misbehaviour, do, however, frequently result in reincarceration. The new offence outcome captured any new crimes committed by JIPs while being supervised in the community. Technical violations and new offences were amalgamated into the any return outcome. Notably, the any return outcome should be understood as any event that *could* result in a return to custody. As explained above, technical violations do not always result in reincarceration, though serious, or repeated violations, do. The majority of the women in this sample committed some kind of violation (67.1%) and a much smaller proportion (10.6%) had their release revoked for a new offence. Given the disparity in base rates, the any return outcome variable was more strongly affected by the

rate of violations. Though this limits the explanatory power of the variable on its own, it was nevertheless desirable to include a variable that captured all recidivism.

Chapter 5: Study 1 – Examining the Instrument

Purpose

Existing DRAOR studies provide ample evidence to support the conceptualization of the DRAOR as a valid, dynamic risk assessment and case management instrument. DRAOR scores have yielded a consistent relationship with recidivism (especially when recidivistic outcomes are defined more generally) and the DRAOR is considered a dynamic instrument in view of its documented ability to capture intra-individual change from one assessment to the next. The DRAOR has also demonstrated measurement invariance across time for general, predominantly male JIP samples (Chadwick, 2014; Hanby, 2013, Lloyd, Hanson, Richards & Serin, 2020), suggesting that differences in first and last assessment scores can reliably be viewed as representing actual change in these JIPs. However, comparable research with women JIPs is lacking and much of research concerning the factor structure of the DRAOR in different subpopulations has yielded inconsistent findings; thus, further research is merited.

Using data from a new, large, sample of JI women on community supervision orders in Iowa, this study examined the factor structure of the DRAOR in a women-only sample. Consequently, this sample differs significantly from samples previously used to examine factor structure (i.e., men from New Zealand and Iowa) where important structural inconsistencies emerged. Competing models were tested in order to determine which provided the best fit for the data. This study also investigated whether the emergent factor structure of the DRAOR was consistent (read invariant) across racial

groups and across time (i.e., between first and last assessment). Without first establishing measurement invariance, it is inappropriate to make comparisons across groups and across assessments (e.g., mean scores, predictive ability of individual items, DRAOR Total and subscale scores, etc.).

To summarize, the overarching purpose of Study 1 was to determine whether or not the DRAOR demonstrated acceptable psychometric properties and if the original three-factor structure would provide a good fit for the DRAOR data at baseline and last assessment for women of different races. Three hypotheses were formulated in response to these questions.

Hypothesis 1. It was expected that the DRAOR would demonstrate adequate psychometric properties, although the original factor structure (as outlined by Serin, 2007) was unlikely to be empirically supported. Alpha values of $\alpha \geq .80$ for the Stable and Protective subscales, and $\alpha \geq .60$ for the Acute subscale, as well as inter-item correlations $r \geq .30$ and concurrent validity with measures of static risk $r \geq .20$ were expected.

Hypothesis 2. It was hypothesized that the DRAOR would meet the criteria for strong factorial invariance (the level required to make comparisons) across racial groups. Findings from previous investigations of the DRAOR have demonstrated that it can be used to accurately predict community outcome for both male and female JIPs (e.g., Hanby, 2013; Yesberg et al., 2015), and for JIPs with different ethnic backgrounds (e.g., Hanby, 2013; Yesberg & Polaschek, 2015), thus, it was expected that the DRAOR would demonstrate sufficient measurement invariance across groups.

Hypothesis 3. The final hypothesis predicted that the DRAOR would demonstrate adequate longitudinal measurement invariance (i.e., strong factorial invariance). Existing research on the DRAOR (Hanby, 2013; Yesberg & Polaschek, 2015) has explored longitudinal measurement invariance. In both cases, the authors conducted Confirmatory Factor Analyses (CFA)s on baseline and subsequent reassessment scores and concluded that there were no significant differences in terms of model fit to their data across these timepoints. These findings were expected to replicate.

Method

Participants

A sample of 3,091 JI women serving community supervision orders in Iowa was utilized in this study. The vast majority of the women in the sample were White (78.2%, n = 2,418). The sample also included 560 Black women (18.1%) and 113 Hispanic women (3.7%). Demographic information (age, marital status, educational achievement and risk level) for these women is presented at the beginning of Chapter 4. Five hundred of these women were subsequently selected for initial exploratory analyses. Note that women were randomly selected within each race group to ensure that sample proportions were maintained.

Procedure and Measures

For a detailed description of the data collection procedure and measures relevant to this study, see Chapter 4.

Analytic Approach

First, the data were restructured using SPSS 25.0 to facilitate data screening and organization prior to primary analyses. Mplus 8.4 (Muthén & Muthén, 1998-2020) was

used for all analyses related to factor structure as well as for supplemental item-level analyses.

Basic psychometrics. A series of simple analyses were conducted first in order to establish whether the DRAOR, in its original format, demonstrates adequate psychometric properties with the current data. Cronbach's coefficient alpha (α), which is the most frequently reported measure of internal consistency was used to assess the overall reliability of the DRAOR and its subscales. Ranging from 0 to 1, larger α values indicate of a greater degree of internal consistency. Though there remains some disagreement over what level of Cronbach's α is required to demonstrate reliability, general guidelines have been established. Reynolds and Livingstone (2013) indicate that coefficients ≥ .70 are likely adequate for this kind of testing situation. This is consistent with Cortina (1993), and Nunnally and Bernstein (1994), who also specified .70 as a minimum cut-off for determining internal consistency. Inter-item correlations and the corresponding score distributions were also examined and assessed in relation to the $r \geq .30$ benchmark suggested by Field (2000) and Schmitt (1996). Finally, DRAOR Total scores were correlated with IVVI scores in order to assess convergent validity.

Exploring the factor structure. Given the size and relatively unique characteristics of the sample, the present study offered an opportunity to perform a robust investigation of the factor structure of the DRAOR in a sample that was not predominately comprised of White men. General convention stipulates that sample sizes greater than 500 cases and case-to-item ratios greater than 10:1 are desirable for factor analysis (Comrey & Lee, 1992; Tabachnick & Fidell, 2013), requirements that were met when considering the overall sample. The present study undertook a systematic approach

to evaluating the factor structure of the DRAOR with the expressed intent of clarifying how individual DRAOR items are related to latent constructs (the subscales) and to inform its application with female correctional populations.

Given the lack of agreement in the research as to the DRAOR's factor structure, it seemed prudent to start at the beginning – that is – with exploratory analyses. Briefly, exploratory factor analysis (EFA) is used to identify underlying latent constructs, and to explore the structure of correlations among variables of interest (Fabrigar et al., 1999). In other words, EFA helps organize data by grouping correlated variables together into overarching sets of factors (each of which is relatively independent of the others; Tabachnick & Fidell, 2013). EFA was selected over principal components analysis (PCA) as PCA is better suited for use when there is no a priori theory underlying how variables are hypothesized to load onto factors (Fabrigar, Wegener, MacCallum, and Strahan, 1999). Moreover, EFA tends to produce a more accurate model of how correlations are structured as it takes error variance into account.

Importantly, EFA analyses relied on a polychloric correlation matrix as opposed to a Pearson Product-Moment correlation matrix. The latter is typically employed in the social sciences but is not appropriate for use with ordinal data as the reduced variability that results from ordinal levels of measurement is problematic for this type of correlation. Numerous researchers (eg., Brown, 2006; Flora & Curran, 2004; Kline, 2016; LaBrish, 2011; Little, 2013) have noted that it often leads to biased estimates and underestimated factor-loadings. Polychloric correlation matrices are not affected by the same issues; simulation studies (e.g., Holgado-Tello, Chacon-Moscoso, Barbero-Garcia, & Vila-Abad,

2010) have demonstrated that polychloric correlations are superior correlation estimators when using ordinal data.

Factor extraction and retention. In accordance with the recommendations of Muthén and Muthén (2010) factors were extracted using a weighted least square (WLSMV) approach. Similarly, a rotation method suitable for use with oblique solutions (promax) was employed as the factors in the model were correlated (i.e., not orthogonal) as per the recommendation of Matsunaga (2010). Several different approaches have been proposed for determining which factors should be retained. These include analyzing scree plots, retaining factors with Eigenvalues greater than 1 (i.e., the Kaiser criterion), and parallel analysis. Despite widespread use, the Kaiser criterion has been criticized by some (e.g., Costello & Osborne, 2005) on the basis that it often leads to inaccurate results and common convention now stipulates that several indices be considered. Accordingly, all three approaches were used in the current study and results were compared.

Factor structure fit. Once plausible models had been identified, the overall fit of each model was examined. Note that the fit indices described below are appropriate for EFA, CFA, and structural equation modeling (ESEM) models and will be used throughout this study. The most well-known and commonly used test to evaluate global model fit is the chi-square test (Cochran, 1952). However, given its well-documented sensitivity to sample size (i.e., it rejects well-fitting models if sample size is large; e.g., Chen, 2007; Cheung & Rensvold, 2002; Kim, Yoon & Lee, 2012; Van den Schoot, Lugtig, & Hox, 2012), other fit indices were also considered. The following fit statistics and corresponding cut-offs were utilized: (a) the Root Mean Square Error of Approximation (RMSEA; cut-off value of < .08 for good fit and ideal < .06), (b) the

Comparative Fit Index (CFI; cut-off value of > .95 for good fit), (c) the Tucker-Lewis Index (TLI; cut-off value of > .95 for good fit), and (d) the Standardized Root Mean Square Residual (SRMR; cut-off value of < .10 for good fit, ideally < .08; Brown, 2014; Hu & Bentler, 1999; Kline, 2005; Marsh, Hau, & Wen, 2004; Matsunaga, 2010; Tsigilis, Gregoriadis, Grammatikopoulos & Zachopoulou, 2018). Notably, fit indices for EFAs based on polychloric correlation matrices identified the correct factor solution more reliably than EFAs based on Pearson correlations (LaBrish, 2011).

Factor loadings. Each item's factor loading was examined to determine which items could accurately be considered indicators for each item. General convention dictates that in order for a variable to load onto a given factor, the factor loadings must meet or exceed .30 (Tabachnick & Fidell, 2013). Indicators that met this criterion were retained and the fit of the resultant three- and four-factor models (in addition to the original DRAOR model) was examined in subsequent steps.

Confirming the factor structure. In CFA, the underlying structure of an instrument (e.g., which items load onto which latent factors) is defined a priori on the basis of theoretical assumptions and/or on the results of earlier EFAs (Tsigilis, et al., 2018). In other words, CFA is a theory-driven approach in which individual items can only be related to one latent construct, and loadings onto all other factors are constrained to zero. This approach has certain advantages. First, substantive a priori knowledge about how items and factors are related can contribute to better, theoretically grounded definitions of the latent variables, thus facilitating interpretation, and second, CFA typically leads to more parsimonious models (Asparouhov & Muthén, 2009). However, there are also disadvantages to the use of CFA measurement models. Despite intuitive

appeal, CFA is predicated on restrictive assumptions which are not always realistic in practice. Assessment instruments frequently have small cross-loadings that are supported by theory or simply a function of how the instrument was developed. In these cases, constraining many or all of these cross-loadings to zero can result in erroneously parsimonious models that are overly simplistic for the data (Asparouhov & Muthén, 2009). Additionally, misspecification of zero loadings in CFA can result in distorted factors and structural relations. If existing cross-loadings are constrained to zero, the correlations between factor indicators and the different factors they intrinsically load onto are forced through their main factor only which can result in inflated factor correlations and misleading factor structures. Although there is a general preference among researchers to seek a confirmatory model, in cases of model uncertainty, exploratory analyses should be considered. As noted by Browne (2001), when a CFA model is rejected, it is common for researchers to undertake a series of model modifications in order to attempt to improve fit. By definition, this process of modification and model refinement becomes exploratory. In this situation, it is arguably preferable to resolve fit issues such as mis-specified loadings by revisiting the factor matrix and rotation rather than by making a series of post hoc model modifications (Browne, 2001; Asparouhov & Muthén, 2009).

Exploratory structural equation modeling is an extension of structural equation modeling that expands the types of measurement models that can be used (Asparouhov & Muthén, 2009; Marsh, Morin, Parker, & Kaur, 2014). The primary advantage of ESEM over other modeling approaches is that it integrates the advantages of EFA and SEM models into one highly flexible modeling approach (Tóth-Király, Bõthe, Rigó, & Orosz,

2017). Specifically, it can estimate structural and measurement models, which is typically accomplished through EFA and subsequent CFA, simultaneously. This avoids the issues often encountered (e.g., poor fit of the CFA model) when moving between EFA and CFA. Importantly, in applied contexts, it is rare for all items in an instrument to load onto one factor only; there is typically at least some degree of relation between items and the non-target factors, thus constraining non-zero cross-loadings to zero can distort the correlations of these items with the one factor they are permitted to load onto (Morin et al., 2016). Furthermore, recent research by Asparouhov, Muthén, and Morin (2015) has demonstrated that nullifying small cross-loadings (as small as 0.100) can result in inflated and biased parameter estimates. Thus, one can argue that in many cases, finding an accurate measurement model should be prioritized over finding a pure set of measurements (Asparouhov & Muthén, 2009). That said, the flexibility of ESEM models comes at the cost of parsimony and interpretability. Asparouhov and Muthén (2009) are also careful to emphasize that researchers need to weigh the benefits of measurement accuracy against their ability to reconcile to model with theory and to explain the resultant structure. A model that provides good fit but makes no theoretical or conceptual sense is likewise undesirable. To summarize, in cases where factor structure is ambiguous following EFA, ESEM *may* be the most advantageous approach. It is considerably more flexible than CFA in that it can estimate measurement and structural models separately but simultaneously, and is likely to give a more accurate, if less parsimonious model.

Given the objectives of this dissertation, a series of CFA and ESEM models were constructed so that the advantages and disadvantages of each type of model could be evaluated. Previous researchers (e.g., Chadwick, 2014, Hanby, 2013, Lloyd, 2015, and

Yesberg & Polaschek, 2015) have tested and ultimately found support for various CFA models with DRAOR data but the variability in the factor structures they identified at the very least speaks to some instability in the DRAOR's structure across correctional subpopulations. The strengths and weaknesses afforded by each model were examined, and the most appropriate model was retained for subsequent invariance testing.

Assessing measurement invariance (MI).

Statistical approaches. After selecting the most advantageous model in terms of theoretical parsimony and model fit, consideration shifted measurement invariance (MI) testing. In other words, did a measure developed based on general desistance literature and theory apply equally well to female JIPs from different racial backgrounds and was this factor structure consistent at the time of the baseline assessment and last assessment? Measurement invariance refers to whether scores from the operationalization of a given construct have the same meaning under different conditions (i.e., across JIP subgroups or time; Kline, 2005; Meade & Lautenschlager, 2004). Importantly, the absence of MI would mean that individuals' scores on the DRAOR could not be reliably compared or unambiguously interpreted as having the same implications in terms of level of risk because differences in DRAOR scores could not be isolated from differences owing to subgroup membership (Horn & McArdle, 1992). Invariance testing was conducted on both the original DRAOR model and on the model that provided the best fit for the present data for comparison purposes.

Multi-group CFA (MGCFA) was used to evaluate MI across racial groups and over assessment time. Multiple indicator multiple cause (MIMIC) modeling was briefly considered but was quickly discounted given that it is a test of uniform MI, which means

that factor loadings are assumed to be invariant across groups and are not explicitly tested for invariance themselves (Woods, & Grimm, 2011). As noted by Kim, Yoon, and Lee (2012), the results of MIMIC modeling are unreliable when the assumption of factor loading invariance is violated, which it often is in practice. Accordingly, MIMIC modeling was eliminated from further consideration.

Criteria for establishing MI. MI is not an all-or-nothing property and can instead be conceived of as a continuum of increasingly stringent hypotheses about invariance, with each successive type of invariance requiring more evidence than the preceding type (Kline, 2005; Wu, Li, & Zumbo, 2007). Importantly, the level of MI supported by the data prescribes the type of inferences that can be drawn. When working with categorical indicators, the basic levels of invariance can be described as equal form (configural) invariance, equal loadings (weak invariance), equal thresholds (strong invariance) and equal residual invariance (strict invariance; Li, Gooden & Toland, 2016; Tsigilis et al., 2018; van den Schoot et al., 2012). Notably, authors frequently use different labels when referring to the same type of invariance (Kline, 2005). Configural invariance simply requires that the number of latent factors is consistent across groups. This type of invariance is tested using EFA and has therefore already been established at this juncture. Weak, or metric, invariance involves constraining only the factor loadings across groups (van den Schoot et al., 2012). Thus, obtaining this type of MI would allow for factor scores (DRAOR domain scores) to be calculated according to the same weighting scheme (Kline, 2005), but would not support the comparison of observed variances or covariances over groups (Asparouhov & Muthen, 2009; Gregorich, 2006). Given the aims of the present project, equal thresholds, or strong invariance is required. This level

of invariance would indicate that both factor loadings and thresholds are equal across groups and would allow unambiguous comparison of DRAOR subscale (therefore also DRAOR Total) scores across the groups being considered. Notably, analyses were conducted separately for MI across race groups and MI across time – the two were examined independently as it is possible to achieve different levels of MI in each case. Strict MI is the most restrictive level of MI and achieving this level of invariance is neither expected nor explicitly required. Equal residual variances would indicate that the latent construct (the DRAOR domains) are measured *identically* across the groups (van de Schoot, Lugtig, & Hox, 2012) which is not required to compare means and variances across groups.

In view of the foregoing, a series of MGCFA models were estimated, and the fit of each increasingly restrictive model was assessed in relation to the previous, less-constrained one (Asparouhov & Muthén, 2009; Cheung & Rensvold, 2002; Kline 2005; Tabachnick & Fidell, 2013). Typically, analyses cease once the addition of more restrictive constraints produces a significant depreciation in model fit, as assessed by change in chi-square value and other model fit indices such as the CFI (Cheung & Rensvold, 2002). Notably, these conventions are applicable for model comparisons when using maximum likelihood estimation and are not appropriate here. The present project used the WLSMV estimator, which does not assume normally distributed variables and is preferred for modelling categorical data (Brown, 2006). Importantly, with WLSMV, chi-square and related fit statistics like CFI cannot be compared across nested models and difference testing can only be carried out in Mplus by using the DIFFTEST option (Muthén & Muthén, 1998-2020). As per the Mplus User's Guide (2010), the DIFFTEST

option is used to obtain accurate chi-square difference test results when using the WLSMV estimator because the difference in chi-square values for nested models is not distributed as a regular chi-square would be. Instead, this test compares the H0 (null) analysis model to a less restrictive alternative model (H1) in which the null model is nested. Obtaining the correct chi-square difference test involves a series of steps. First, the H1 model is estimated and the derivatives required for model difference testing are saved. Next, the H0 model is estimated, and the derivatives from this model are saved. Finally, the chi-difference test is computed using the derivatives for the H0 and H1 analyses. Notably, when comparing chi-square values obtained from models using the ML estimator, large sample sizes can be problematic as they yield a very liberal significance test (i.e., the null is almost always rejected; Cheung & Rensvold, 2002); this is not an issue when interpreting the significance of the chi-square value when using the DIFFTEST procedure as the number of parameters being compared and the degrees of freedom for the H0 – H1 comparison are much smaller.

Supplementary analyses. Differential item functioning (DIF) analyses were performed to attempt to better understand the results that emerged following the measurement invariance analyses. Briefly, DIF arises when an indicator has a meaningfully different pattern of coefficients or intercepts over groups (Kline, 2005), meaning that an individual's score on the indicator, given their true score on the corresponding factor, will be influenced by subgroup membership. Importantly, the presence of DIF is not necessarily indicative of bias (Walker, 2011), but does suggest the need for further investigation. DIF procedures are based on item response theory (IRT) and can be employed to gain a more nuanced understanding of the performance of

individual items within an overall measure (Simms, 2008; Walker, 2011). Interested readers are encouraged to refer to Embretson and Reise (2000) for a more in-depth treatment of IRT, and to Walker (2011) for an overview of DIF. The goal of these analyses was to identify any items that might be responsible for producing large differences. Of note, one of the assumptions underlying simple IRT is that the measure be unidimensional; as multidimensionality and basic IRT are incompatible, a more complex variation (multidimensional IRT) was necessary here (Embretson & Reise, 2000; Van Dam, Earlywine, & Borders, 2010). Category response curves (CRCs) and item information curves (IICs) were used to examine the discriminative abilities and amount of information yielded by individual items. Briefly, CRCs represent the probability that a JIP with a given trait level (θ) will be assessed as belonging to a particular category (i.e., scored 0, 1, or 2 on the item) and IICs depict the amount of information that each item yields on its own and in proportion to the total scale (Van Dam et al., 2010). Considered together, CRCs and IICs are a useful metric for determining how much individual items contribute to the scale's utility. The utility of individual DRAOR items was assessed and implications are discussed.

Results

Data Management and Assumptions Tests[9]

DRAOR items and subscale scores were screened to ensure that they were in the valid range. No issues were found. Also, as cases with missing data on key variables (i.e.,

[9] The majority of data management and screening was completed in preparation for Study 1 and will not be revisited in subsequent studies. Only the results of assumptions checks that were not required for the present study will be described in Studies 2 and 3.

race) were excluded from analyses earlier, no additional issues with missing data were found.

Basic assumptions were examined across DRAOR Total and subscale scores. Specifically, linearity, homoscedasticity, normality, univariate and multivariate outliers, and multicollinearity were assessed. Inspection of standardized *z*-scores and corresponding box plots indicated that univariate outliers were not an issue and bivariate scatterplots suggested that linear relationships existed for each combination of DRAOR subscales. Equally, the assumption of homoscedasticity was supported, as variability in scores for each variable was consistent across all values of the other variables. As per the recommendation of Tabachnick and Fidell (2013), normality was assessed by visually inspecting the shape of the distributions as opposed to relying on *z*-scores and standard tests of normality. When sample size is large, Kolmogorov-Smirnov's test of normality often identifies departures from normality and researchers should adjust their concept of what counts as an outlier. When sample size is in the thousands, relying on conventional wisdom (i.e., the +/- 3 standard deviations rule) is too conservative; with samples this large, it is reasonable to have some cases scoring 3 and 4 standard deviations from the mean. In such cases, it is more useful to reserve the designation of "outlier" for cases that are qualitatively different from others in the sample and that may bias results rather than cases with scores that are only slightly higher than those below them but haven't crossed the pre-determined cut-off value. Accordingly, histograms, probability plots, and skewness and kurtosis ratios were examined, and these indicated only minor departures from normality. Also, inspection of the graphed data suggested a normal distribution. Mahalanobis' distance was used to test for the presence of multivariate outliers. One

problematic case was identified. However, subsequent consideration of influence values (i.e., Cook's distance) indicated that the case did not qualify as an outlier based on this index (Cook's $D < 1.00$) and the case was retained. Finally, bivariate correlations were used to test for singularity and multicollinearity. No issues with multicollinearity were identified, as correlations between the DRAOR subscales did not suggest a perfect, or near perfect relationship. The strongest relationship was between the Stable and Acute risk scales; Table 4 provides the inter-correlation values.

Table 4

Inter-Correlations Between Original DRAOR Subscales

	Acute Risk	Protective
Stable Risk	.615**	-.508**
Acute Risk		-.469**

** p <.01

Psychometrics and Factor Structure of the Original DRAOR

Initial exploratory analyses were conducted on a small, randomly selected subset of the overall dataset (n = 500) prior to assessing the fit of CFA models with the larger portion of the sample. There were no significant differences between women in the subsample and the main sample with regards to age, race, education, and marital status (see Appendix C). A sample size of 500 was chosen as it was sufficiently large as to be adequate for conducting an exploratory factor analysis while retaining the bulk of the sample for later confirmatory analyses. Although roughly half of the communalities were below .5, a sample size of 500 can reasonably be expected to produce reliable factors (MacCallum, Widaman, Zhang, & Hong, 1999). Additionally, a Kaiser-Meyer Olkin measure of sampling adequacy of 0.89 suggests that the data are adequate for factor analysis. The subject to variable ratio of 26:1 is also acceptable.

A polychloric correlation matrix consisting of the DRAOR's 19 items cross-correlated with each other was used to further examine the factorability of the data and to assess whether singularity or multicollinearity issues were present in the data at the item-level. Thirty eight percent of the cross-correlations (total = 171) were .30 or greater, with several exceeding .50. Barlett's test of sphericity was significant (χ^2 (*df* = 171) = 2920.97, $p < .001$) indicating that the **R** matrix was factorable. Moreover, the determinant of the **R** matrix (.003) did not suggest issues of multicollinearity (Tabachnick & Fidell, 2013). Considered collectively, these findings indicate that factor analysis was justifiable.

The results of the initial EFAs did not provide support the original three-factor structure and instead suggested two alternative models; a different three-factor model and four-factor model. Importantly, the appropriate number of factors to retain differed according to which retention criteria were used. Examination of the scree plot and Kaiser's criterion (i.e., eigenvalues > 1.00) recommended that four factors be retained. However, the accuracy of the Kaiser criterion has been criticized by some (e.g., Costello & Osborne, 2010, Field, 2005) who suggest that it often leads to the retention of too many factors. Accordingly, parallel analysis, which compares randomly generated eigenvalues to those obtained in the sample was also conducted to provide a basis for comparison. Note that although the appropriateness of parallel analysis with item-level data has also been criticized, this applies primarily to parallel analysis based on Pearson's *r* correlations. There is considerable evidence that suggests that basing the analysis on polychloric correlations instead yields more reliable results (Cho & Bandalos, 2009; Garrido et al., 2013; Timmerman & Lerenzo-Seva, 2011). This approach suggested that three factors be retained. Based on these findings, three different models (the original

DRAOR and the 3- and 4-factor models suggested by the EFA) were retained for later analyses.

With respect to the psychometric properties of the original DRAOR, the hypothesis suggesting that they would be adequate was not fully supported with the present sample of women. While internal consistency of the three subscales was found to be adequate (Cronbach's α = .76 for the Stable subscale, .67 for the Acute subscale, and .85 for the Protective subscale) and DRAOR scores were correlated with the IVVI (convergent validity, r = .69), many of the inter-item correlations did not reach the $r \geq$ .30 benchmark suggested by Field (2000) and Schmitt (1996). As illustrated in Table 5 below, the Acute subscale was highly problematic. Conversely, the inter-item correlations seen in the Protective subscale are quite strong and those for the Stable subscale are somewhat mixed. Collectively, these results suggest that it is the risk, rather than the protective items that hinder the scale's performance.

Table 5

Inter-Item Correlations for the Stable, Acute, and Protective Subscales of the Original DRAOR

		S1	S2	S3	S4	S5	S6	
Stable	S1	1.00						
	S2	.30	1.00					
	S3	.40	.34	1.00				
	S4	.40	.35	.62	1.00			
	S5	**.18**	.47	**.29**	.31	1.00		
	S6	**.27**	.35	.30	.32	.36	1.00	
		A1	A2	A3	A4	A5	A6	A7
Acute	A1	1.00						
	A2	**.16**	1.00					
	A3	**.08**	**.24**	1.00				
	A4	.33	**.29**	**.18**	1.00			
	A5	**.25**	**.12**	**.11**	**.25**	1.00		
	A6	**.22**	**.19**	**.17**	**.26**	**.21**	1.00	
	A7	**.29**	**.18**	**.16**	**.29**	.31	.38	1.00
		P1	P2	P3	P4	P5	P6	
Protective	P1	1.00						
	P2	.49	1.00					
	P3	.48	.52	1.00				
	P4	.53	.55	.53	1.00			
	P5	.42	.40	.47	.42	1.00		
	P6	.48	.50	.50	.51	.59	1.00	

Note. **Bold** denotes correlations that did not reach the $r \geq .30$ benchmark. S# = Stable item, A# = Acute item, P# = Protective item.

Exploring the factor structure. Tables 6 and 7 present the model fit indices and factor loadings for the alternative three- and four-factor models that emerged following the EFAs. Both models provided adequate fit for the data (χ^2 (117) = 289.17, p < .001 for the three-factor model and χ^2 (101) = 198.36, p < .001 for the four-factor model) though the lower chi-square value associated with the four-factor model suggests slightly better fit. Evaluation of the model fit indices confirms this, with both models exceeding the commonly accepted benchmarks for good model fit (i.e., RMSEA < 0.08, CFI and TLI > 0.95 and SRMR, 0.08; Brown, 2014; see also Hu & Bentler, 1999; Kline, 2005; Marsh, Hau, & Wen, 2004; Matsunaga, 2010; Tsigilis et al., 2018), though again, fit for the four-factor model was marginally better.

Table 6

Model Fit Indices for the Alternative Three- and Four-Factor Models that Emerged Following Exploratory Factor Analyses (EFA) with the Random Subsample (N = 500)

	3 Factor Model	4 Factor Model
χ^2 (*df*)	289.17** (117)	198.36** (101)
RMSEA	0.054	0.044
CFI	0.971	0.983
TLI	0.957	0.972
SRMR	0.046	0.037

Note. χ^2 = chi-square, *df* = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker-Lewis Index, SRMR = standardized root mean square residual. ** denotes significance at the *p* < .001 level.

Table 7

Pattern of Factor Loadings for the Alternative Three- and Four-Factor Models that Emerged Following Exploratory Factor Analyses (EFA) with the Random Subsample (N = 500)

	3 Factor Model			4 Factor Model			
DRAOR Item	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
Peer associations	0.711	-	-	0.596	-	-	-
Attitudes toward authority	-	0.768	-	-	0.755	-	-
Impulse control	0.822	-	-	0.732	-	-	-
Problem solving	0.834	-	-	0.686	-	-	-
Sense of entitlement	-	0.702	-	-	0.691	-	-
Attachment with others	-	0.570	-	-	0.590	-	-
Substance abuse	0.763	-	-	0.576	-	-	-
Anger/hostility	-	0.772	-	-	0.740	-	-
Access to victims	-	0.433	-	-	0.438	-	-
Negative mood	0.602	-	-	0.426	-	-	-
Employment	0.659	-	-	0.426	-	-	-
Interpersonal relationships	0.475	-	-	-	-	0.468	-
Living situation	0.425	-	-	-	-	0.412	-
Responsiveness to advice	-	-	0.599	-	-	-	0.618
Prosocial identity	-	-	0.620	-	-	-	0.658
High expectations	-	-	0.792	-	-	-	0.763
Costs/benefits	-	-	0.610	-	-	-	0.701
Social support	-	-	0.857	-	-	-	0.738
Social control	-	-	0.830	-	-	-	0.741

Notably, the three-factor model obtained following the EFA differs considerably from the DRAOR's current (i.e., original) structure. While the items comprising the protective subscale did load together onto one factor, the items comprising the Stable and Acute subscales did not map onto the factors as expected. The alternative four-factor model was identical to the alternative three-factor model except for two items, interpersonal relationships and living situation, which loaded onto their own factor. Figures 1, 2 and 3 below illustrate the differences between the original DRAOR and the two alternative models.

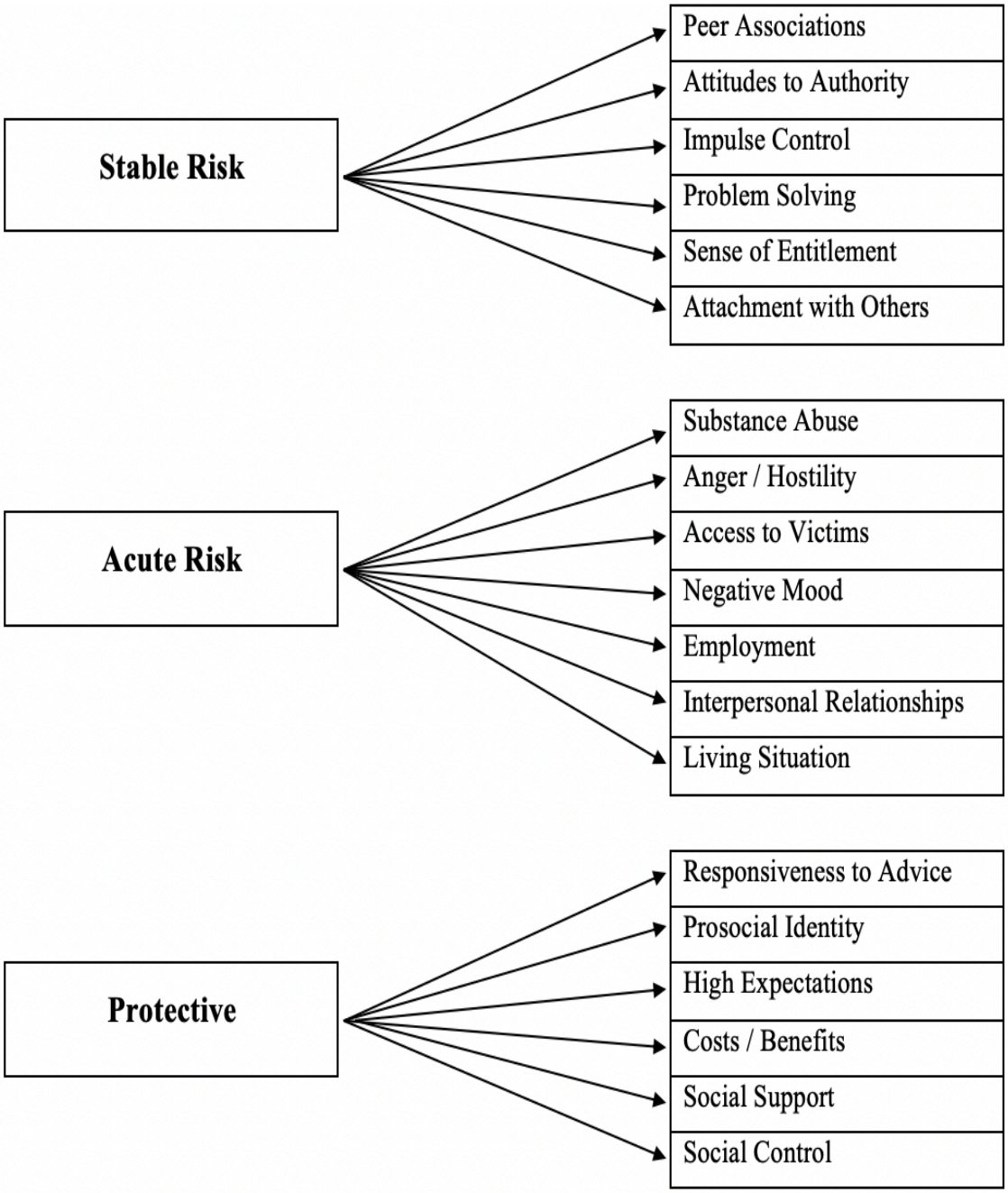


Figure 1. Factor Structure of Original Three-Factor DRAOR.

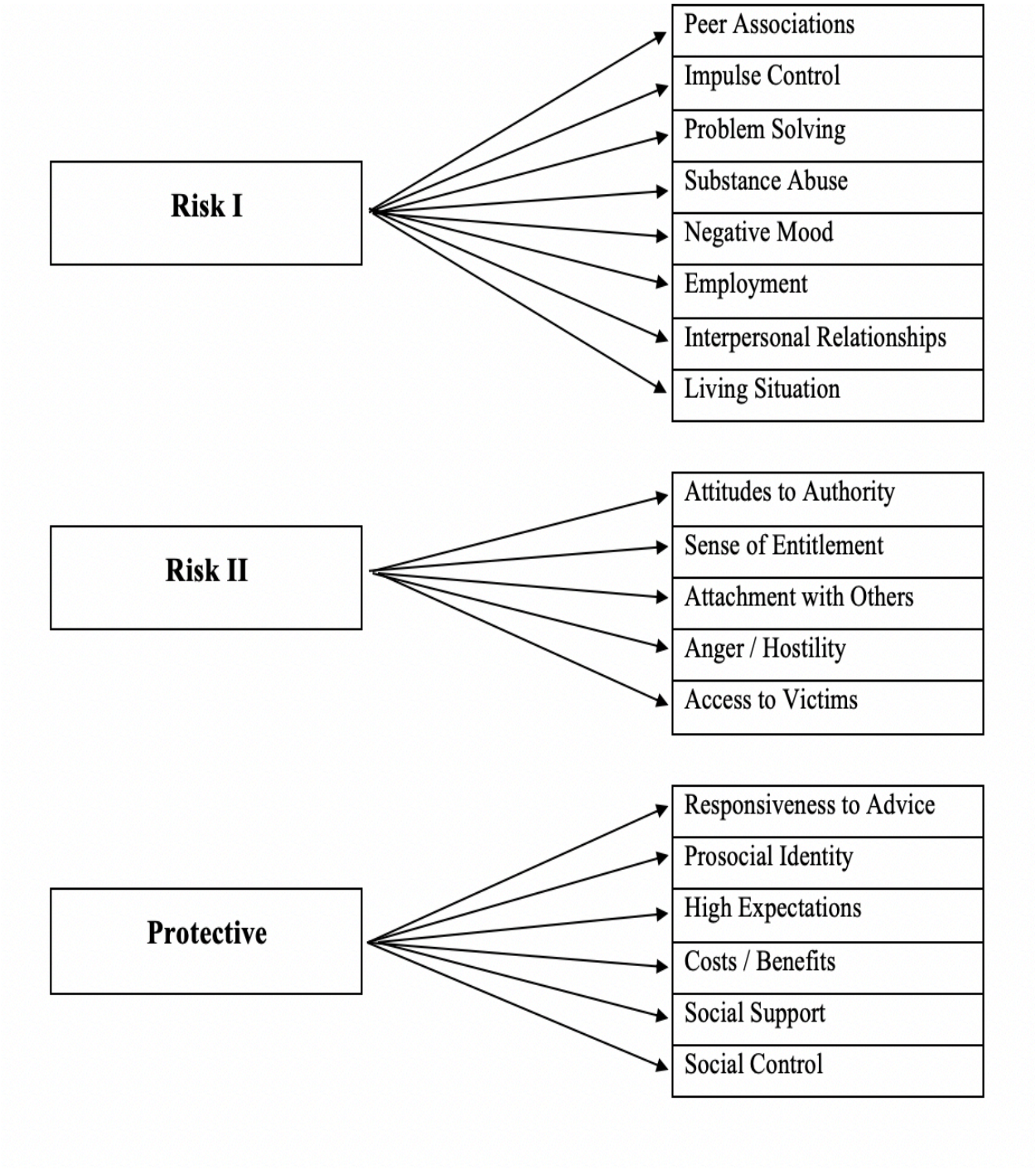


Figure 2. Factor Structure of Alternative Three-Factor Model.

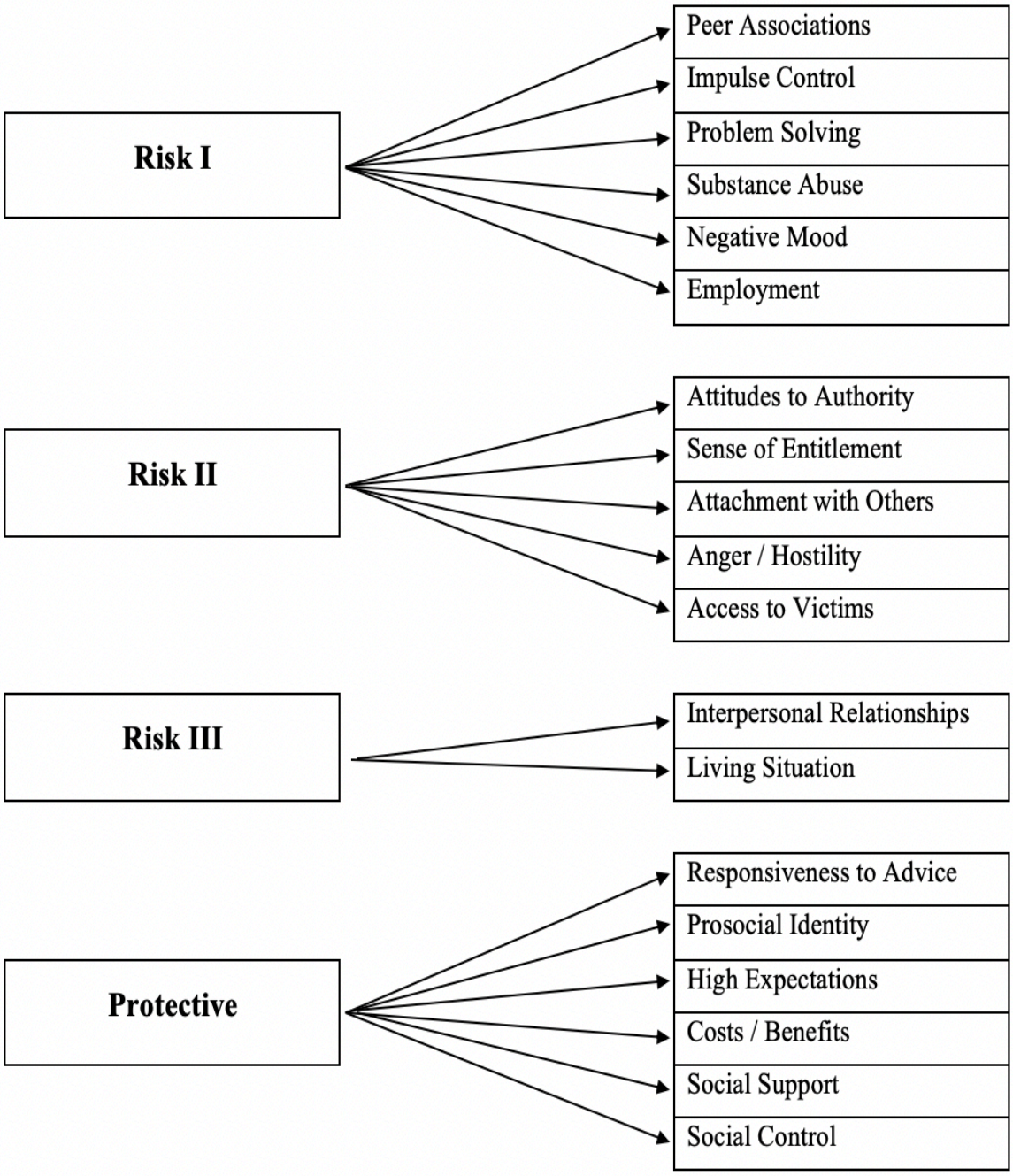


Figure 3. Factor Structure of Alternative Four-Factor Model.

Selecting the best factor structure. As both alternative models fit the data well, both were tested alongside the original DRAOR as CFA models using the larger sample of 2,591 women. Exploratory structural equation modeling (ESEM) techniques were also applied to determine if a more flexible model would provide a better fitting model. Table 8 below summarizes the model fit and factor structure for the original DRAOR.

Table 8

Model Fit Indices and Factor Loadings for the DRAOR with Full Sample (n = 2,591)

χ^2 (*df*)		2637.69** (149)	
RMSEA		0.079	
CFI		0.923	
TLI		0.911	
SRMR		0.058	
DRAOR Item	Stable	Acute	Protective
Peer associations	0.683	-	-
Attitudes toward authority	0.695	-	-
Impulse control	0.804	-	-
Problem solving	0.813	-	-
Sense of entitlement	0.573	-	-
Attachment with others	0.608	-	-
Substance abuse		0.530	-
Anger/hostility	-	0.495	-
Access to victims	-	0.353	-
Negative mood	-	0.513	-
Employment	-	0.471	-
Interpersonal relationships	-	0.533	-
Living situation	-	0.570	-
Responsiveness to advice	-	-	0.784
Prosocial identity	-	-	0.793
High expectations	-	-	0.757
Costs/benefits	-	-	0.803
Social support	-	-	0.711
Social control	-	-	0.813

Note. χ^2 = chi-square, *df* = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker-Lewis Index, SRMR = standardized root mean square residual. ** denotes significance at the $p < .001$ level.

Considered collectively, these results suggest that the DRAOR, in its original form, provide an adequate, but not ideal fit for the data. The RMSEA, CFI, and TLI values can be considered good and the SRMR value is excellent (e.g., Hu & Bentler, 1999; Kline, 2005; Marsh, Hau, & Wen, 2004; Matsunaga, 2010; Tsigilis et al., 2018). However, some of the factor loadings are on the low side. Despite loading significantly onto the factor, a few of the indicators (e.g., access to victims and employment) have weaker loadings. Comparatively, the competing three- and four-factor models suggested by the EFAs provide improved fit. Not only are the model chi-square values smaller, but all of the additional fit indices also suggest improved fit (see Table 9). Moreover, on average, the indicators in the alternative three- and four-factor models load more strongly onto their factors (see Table 10). Ignoring the factor loadings for the six items that comprise the protective subscale (they are equal in all three models), the average loading for the combined indicators in the Stable and Acute risk subscales is 0.59 in the original model. Averaging the loadings of these same indicators in the alternative models results in an average loading strength of 0.65 for both models.

Table 9

Model Fit Indices for the Alternative Three- and Four-Factor Models that Emerged Following Confirmatory Factor Analyses (CFA) with Full Sample (n = 2,591)

	3 Factor Model	4 Factor Model
χ^2 (*df*)	1796.92** (149)	1699.15** (146)
RMSEA	0.064	0.063
CFI	0.949	0.952
TLI	0.941	0.944
SRMR	0.048	0.046

Note. χ^2 = chi-square, *df* = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker-Lewis Index, SRMR = standardized root mean square residual. ** denotes significance at the $p < .001$ level.

Table 10

Pattern of Factor Loadings for the Alternative Three- and Four-Factor Models that Emerged Following Confirmatory Factor Analyses (CFA) with Full Sample (n = 2,591)

	3 Factor Model			4 Factor Model			
DRAOR Item	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
Peer associations	0.842	-	-	0.714	-	-	-
Attitudes toward authority	-	0.812	-	-	0.812	-	-
Impulse control	0.958	-	-	0.826	-	-	-
Problem solving	0.877	-	-	0.838	-	-	-
Sense of entitlement	-	0.660	-	-	0.660	-	-
Attachment with others	-	0.696	-	-	0.695	-	-
Substance abuse	0.529	-	-	0.533	-	-	-
Anger/hostility	-	0.551	-	-	0.551	-	-
Access to victims	-	0.380	-	-	0.380	-	-
Negative mood	0.503	-	-	0.507	-	-	-
Employment	0.468	-	-	0.470	-	-	-
Interpersonal relationships	0.527	-	-	-	-	0.656	-
Living situation	0.564	-	-	-	-	0.707	-
Responsiveness to advice	-	-	0.784	-	-	-	0.784
Prosocial identity	-	-	0.793	-	-	-	0.793
High expectations	-	-	0.757	-	-	-	0.757
Costs/benefits	-	-	0.803	-	-	-	0.803
Social support	-	-	0.711	-	-	-	0.711
Social control	-	-	0.813	-	-	-	0.813

In the interest of accuracy, three- and four-factor ESEMs were also analyzed. Unsurprisingly, as ESEM involves freeing-up parameters so that indicators may cross-load onto multiple factors, both of these models provided a better fit to the data insofar as smaller chi-square values are concerned. That said, there was little practical difference between the corresponding three-factor CFA and ESEM and the four-factor CFA and ESEM with regards to the other indices of model fit. Put simply, all four models fit very well. However, inspection of the cross-loadings produced by these models (see Appendix D) supports the espousal of a CFA model over an ESEM model. With the exception of three indicators in the four-factor ESEM model, all of the indicators loaded most strongly onto the factors specified in the CFA analyses and the majority of cross-loadings were small and non-significant. Of the cross-loadings that were significant, only two exceeded 0.30. Interestingly, in the four-factor ESEM model, the pattern of loadings was not consistent with a four factor model; the two indicators that loaded onto their own factor in the four-factor CFA model (interpersonal relationships and living situation) loaded most strongly onto the factor described by peer associations, impulse control, problem solving, substance use, negative mood, and employment in the three-factor model.

Final model decisions. Table 11 summarizes the fit of the seven models tested so far in this study. After careful consideration, the alternative three-factor CFA model was ultimately selected as the best model. The ESEM models, though marginally better fitting, were less appealing than the CFA models on account of the added complexity in interpretation that they engender. Relatedly, the fact that the factor loadings that emerged from the four-factor ESEM seemed more representative of a three-factor model was concerning. As noted earlier, a model that fits well but makes little conceptual sense is not superior to one that fits slightly less well objectively but it more theoretically appealing (Asparouhov & Muthén, 2015). Of the CFA models, the alternative three-factor model provided the best balance of model fit and theoretical parsimony and was therefore selected as the model used to explore measurement invariance. Although the four-factor model was identified despite one factor being represented by two indicators, having interpersonal relationships and living situation as the sole determinants of a separate factor (construct) was conceptually unappealing.

Table 11

Model Fit Indices for all Models Tested

	3-Factor EFA	4-Factor EFA	Original DRAOR CFA	Alternative 3-Factor CFA	Alternative 4-Factor CFA	Alternative 3-Factor ESEM	Alternative 4-Factor ESEM
χ^2 (*df*)	289.17** (117)	198.36** (101)	2637.69** (149)	1796.92** (149)	1699.15** (146)	1369.88** (117)	795.35** (101)
RMSEA	0.054	0.044	0.079	0.064	0.063	0.063	0.051
CFI	0.971	0.983	0.923	0.949	0.952	0.951	0.978
TLI	0.957	0.972	0.911	0.941	0.944	0.943	0.964
SRMR	0.046	0.037	0.058	0.048	0.046	0.036	0.027

Note. χ^2 = chi-square, *df* = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker-Lewis Index, SRMR = standardized root mean square residual. ** denotes significance at the $p < .001$ level.

Measurement Invariance Testing

Across racial groups. Multigroup CFA (MGCFA) was used to test for measurement invariance in the alternative three-factor model of the DRAOR across the following three racial groups: White women, Black women, and Hispanic women. Configural invariance (i.e., dimensional invariance, or basic structural invariance) had already been established through EFA. Therefore, the next step was to test for weak factorial invariance. Contrary to expectations, significant results for Mplus' chi-square difference testing (DIFFTEST) option indicated that this level of invariance was not supported (χ^2 (48) = 81.67, $p < .01$).[10] While the present study did not expect to find strict invariance across racial groups, this level of non-invariance was unexpected. A consequence of these findings is that no comparisons can be made across racial groups and that it is inappropriate to aggregate women of different racial groups into a larger female sample for analyses.

Over time. To assess longitudinal measurement invariance (i.e., MI over time), MGCFA was used to compare the structure of the DRAOR (the alternative three-factor model) at Time 1 (baseline) and Time 2 (last assessment). Results indicated that weak factorial invariance was achieved (χ^2 (16) = 19.63, $p = .237$) and that the factor loadings were therefore equal at baseline and last assessment. However, the next level of invariance, strong factorial invariance, was not achieved (χ^2 (36) = 123.92, $p < .001$). Again, these results were at odds with expectations; it was hypothesized that the DRAOR would demonstrate sufficient longitudinal measurement invariance to support

[10] Analyses using the original DRAOR structure also resulted in non-invariance across racial groups at the weak factorial level: χ^2 (48) = 91.52, $p < .001$.

comparisons between baseline and later assessment scores. While weak measurement invariance does allow variances and covariances to be compared, means cannot be compared, thus, achieving only this level of MI is of little practical significance. [11]

Investigating Non-Invariance: Supplemental Analyses

Category response curves (CRCs) and item information curves (IICs) were generated in Mplus to better understand how individual DRAOR items contributed to the discriminative and informative functions of the scale. Figure 4 below provides a graphical display of the amount of information that each of the 19 items yields in proportion to the total scale. Fourteen of the 19 items have completely flat IICs and contribute only tiny fractions of information. Items 2, 5, 6, 8, and 9 are the only items that individually contribute a noticeable amount of information (Figure 5), and collectively account for vast majority of the total information provided by the scale (see Figure 6).

[11] Consistent with the alternative three-factor structure, weak factorial invariance between first and last assessment (Time 1 and Time 2) was also achieved using the original DRAOR structure (χ^2 (16) = 12.18, p = 0.73) and strong factorial invariance was not (χ^2 (36) = 119.31, p < .001).

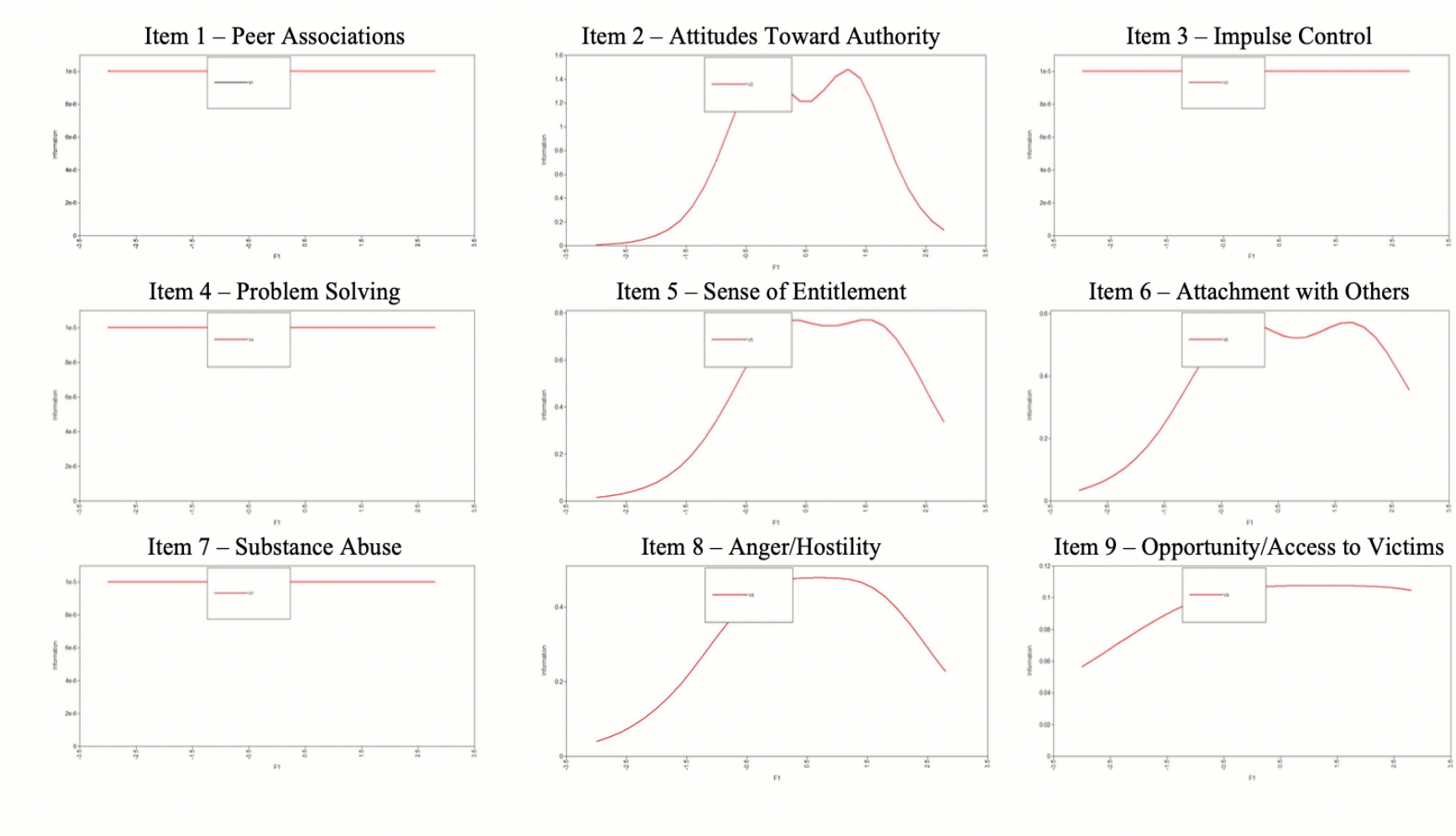


Figure 4. Item Information Curves (IICs) for the DRAOR's 19 Items.

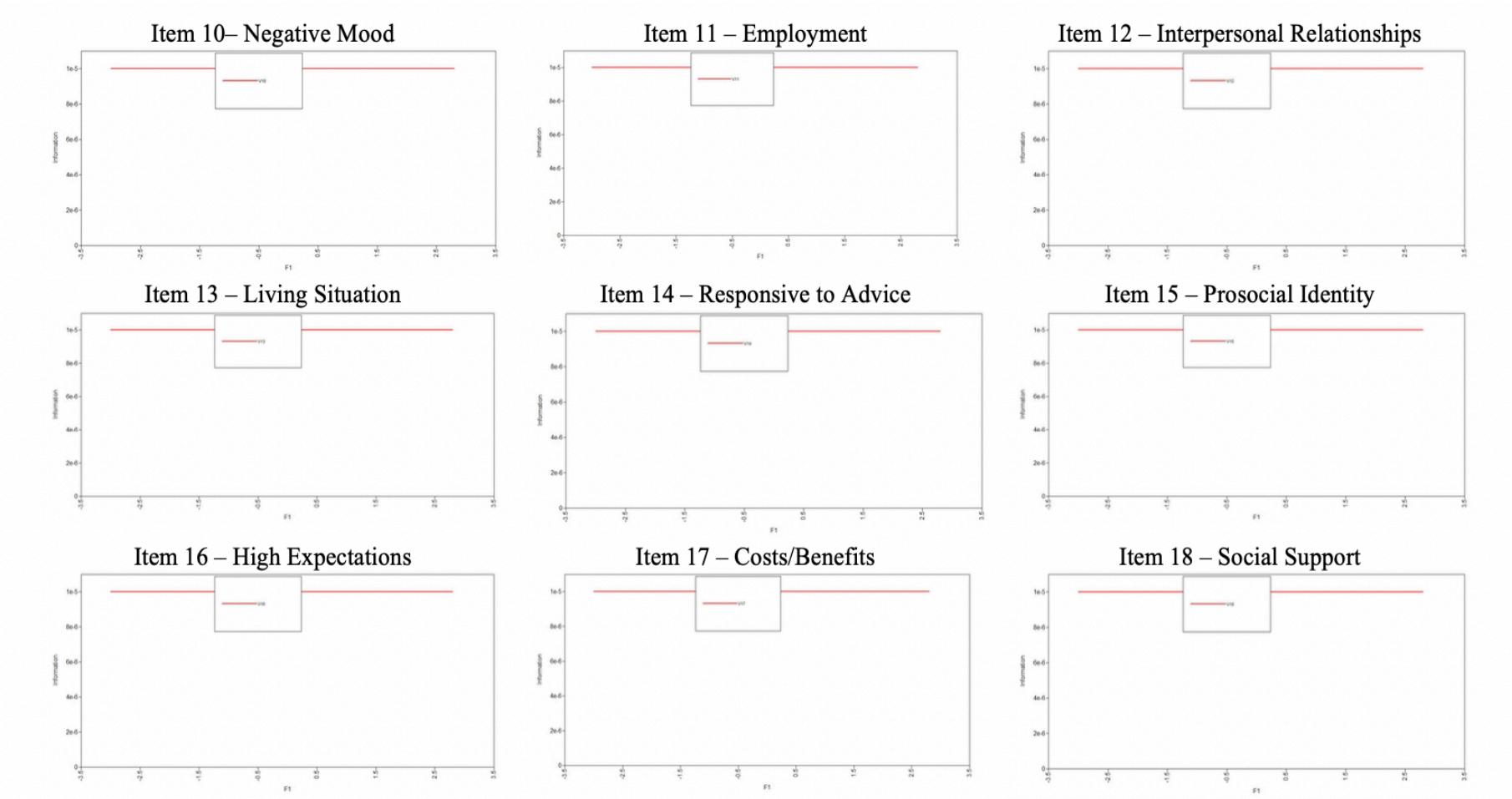
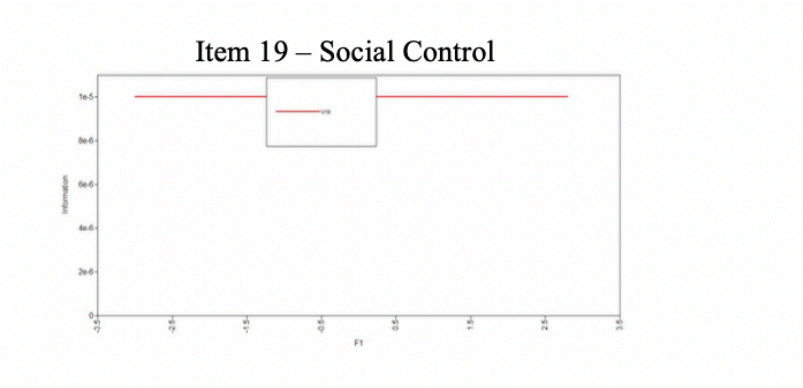


Figure 4. Continued.



Note. #e-# is notation for very small numbers (2e-6 = 2 to the minus 6 power or .000002)

Figure 4. Continued.

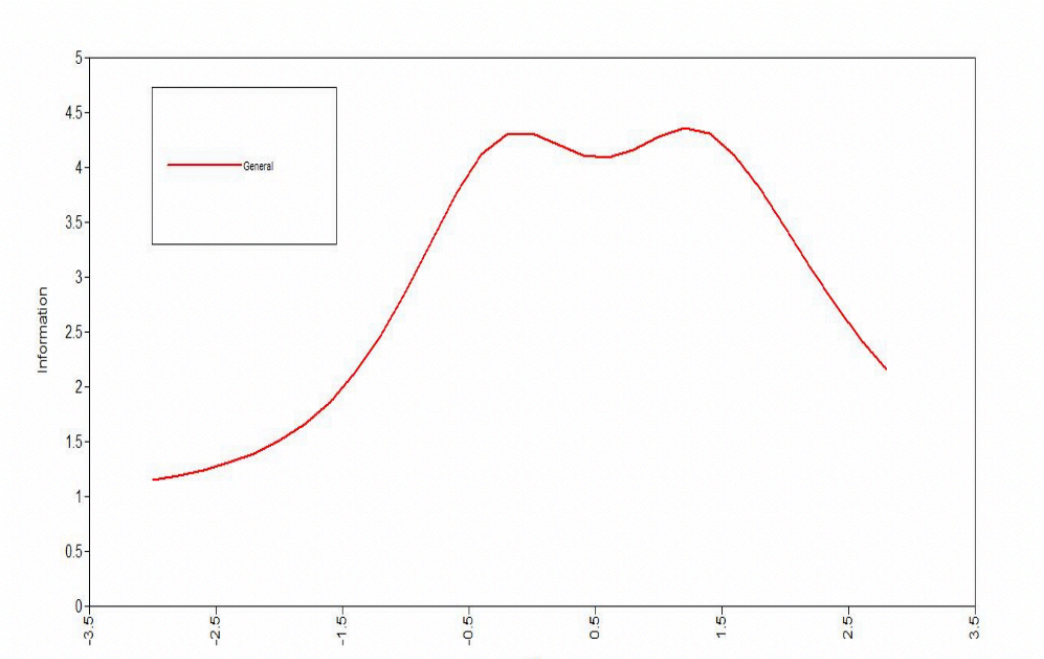


Figure 5. Total Information Provided by DRAOR Items.

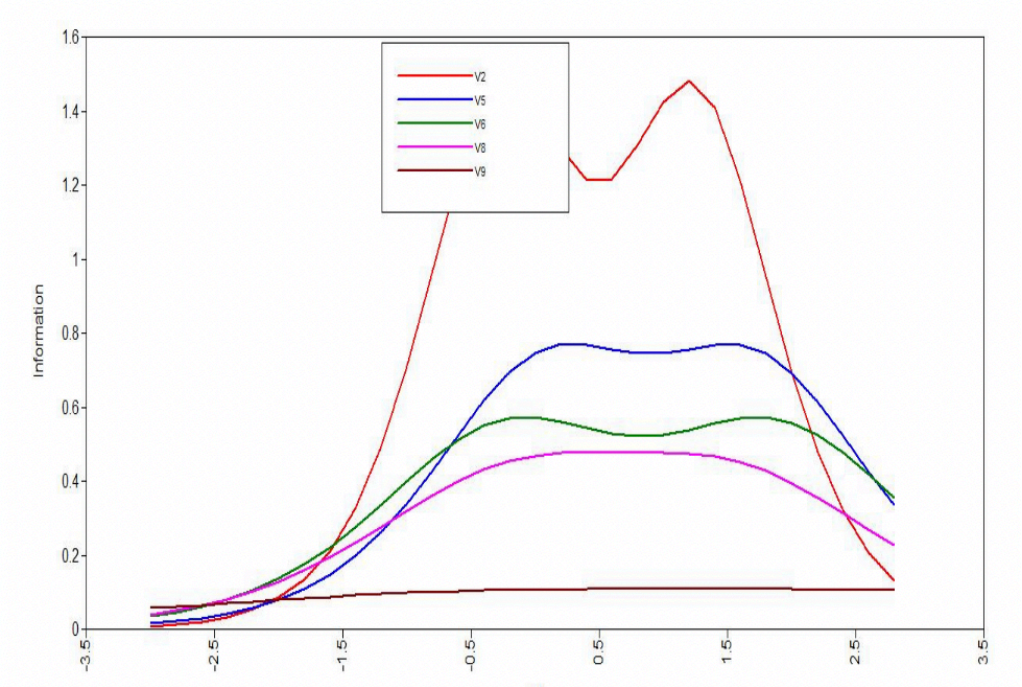


Figure 6. Information Provided by Five Most Informative DRAOR Items.

Simultaneously examining Figures 4 (above) and 7 (below) reveals that low information corresponds to undifferentiating (flat) CRCs. CRCs represent the probability that an individual with a given trait level (i.e., level of risk on a particular item) will be rated as belonging to a particular category (i.e., given a score of 0, 1, or 2). For example, if we consider the CRC for item 2 (attitudes to authority) we can see that it provides good discrimination. The blue curve for category two (i.e., a score of 1 – *slight/possible problem*) covers middle of the horizontal axis and has the highest probability of being endorsed for JIPs displaying a mean level of this trait. The curve for category 1 (a score of 0 – *not a problem*) is highest toward the lowest end of the horizontal axis, thus indicating that JIPs with lower than average need in this area are most likely to receive a score of 0. Conversely, the green curve, representing the probability of receiving a score of 2, is most elevated when levels of the underlying trait (problems in the area of attitudes to authority) are high. Considered collectively, we can conclude that item 2 discriminates well between JIPs who do and do not have problematic attitudes toward authority and that it accounts for a sizeable amount of the total information provided by the scale.

Item 1 (peer associations), on the other hand, appears to contribute relatively little to the overall scale utility. The CRC for this scale suggests that roughly 50% of JIPs are likely to receive a score of 1 (*slight/possible problem*), and that most of the remaining JIPs (40%) will receive a score of 2 (*definite problem*) regardless of the actual severity of their problems with antisocial peers. Even if JIPs do not have problematic peer associations (those to the far left on the horizontal axis), less than 10% of them will receive a score of 0 on this item.

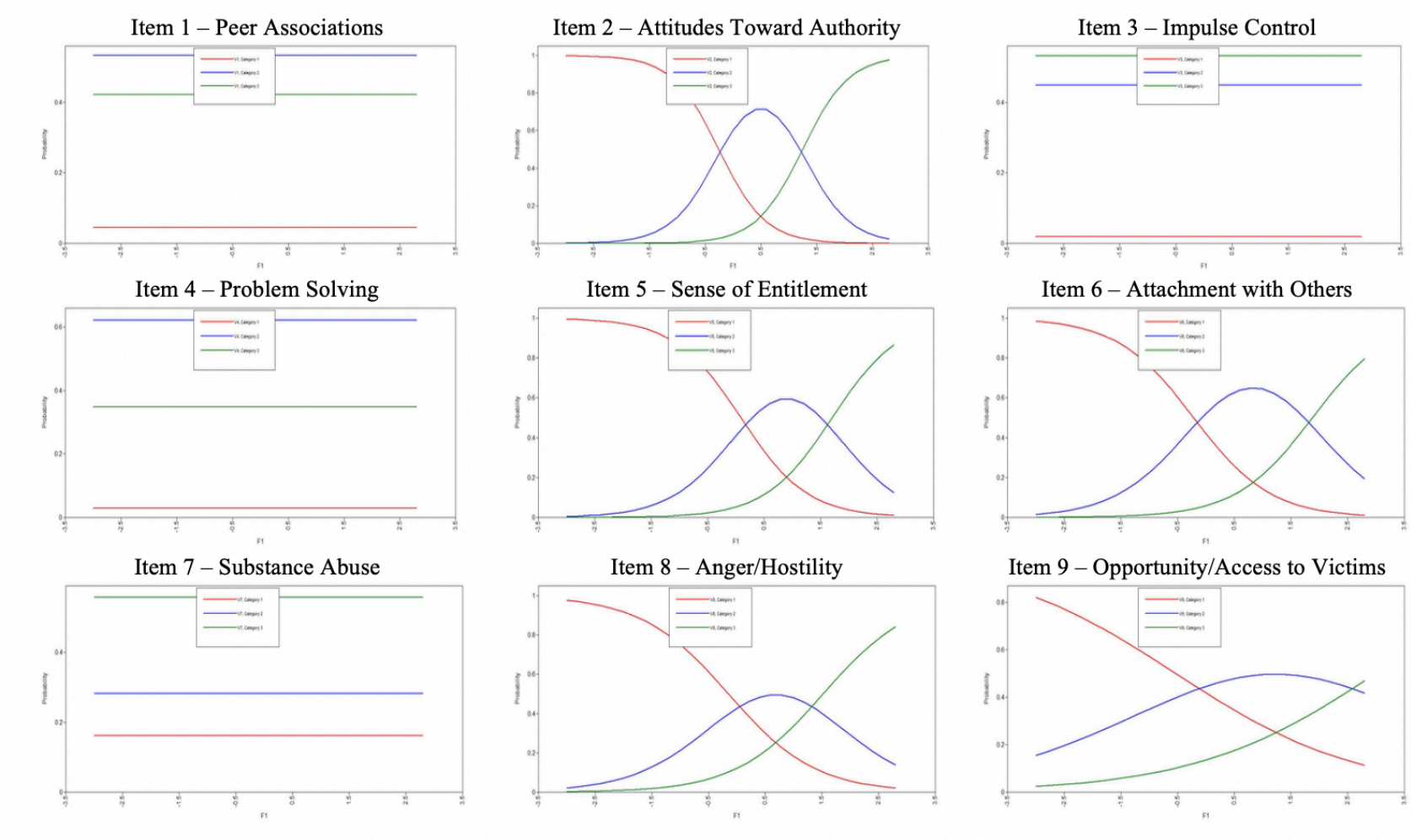


Figure 7. Characteristic Response Curves (CRCs) for DRAOR's 19 Items.

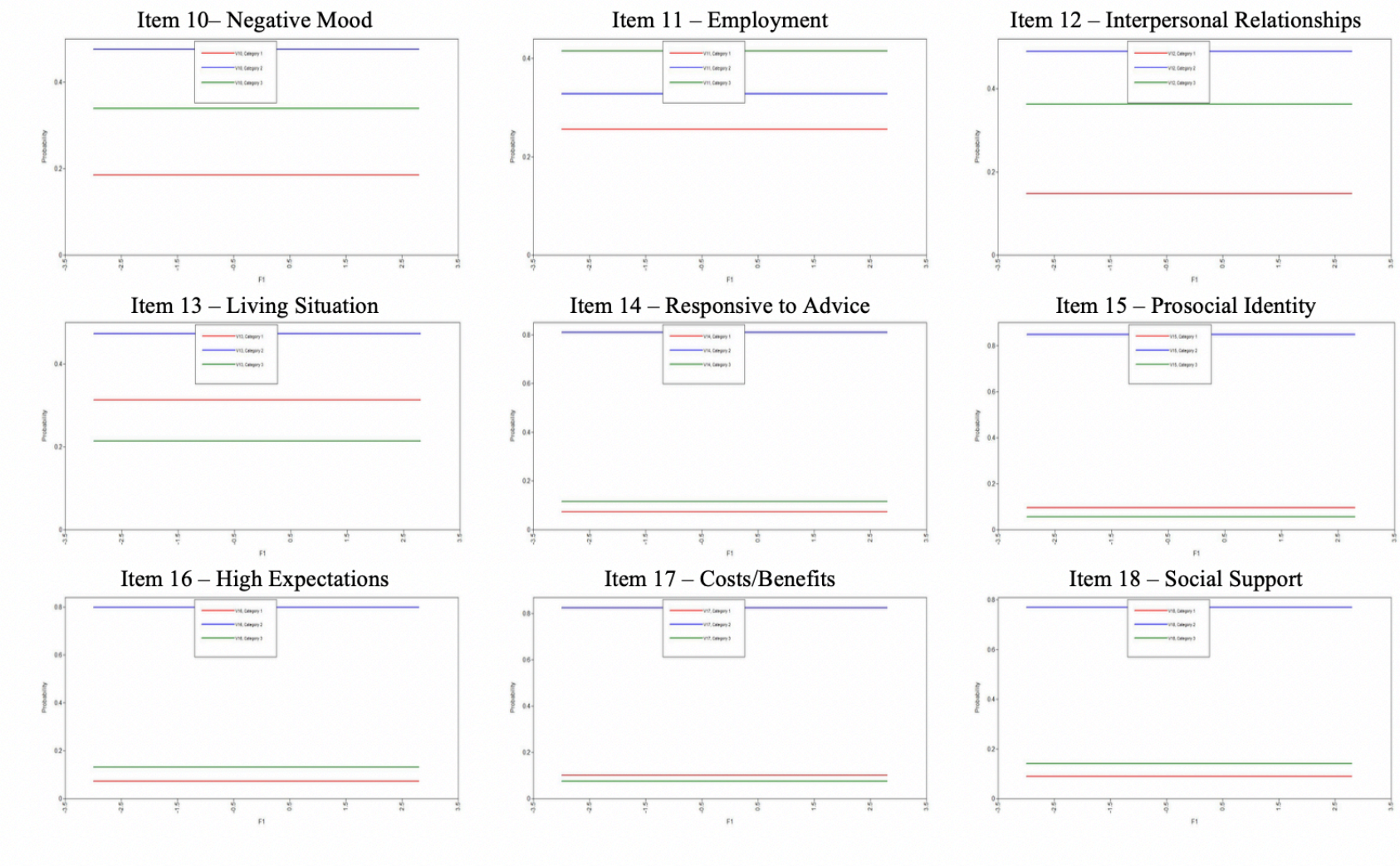
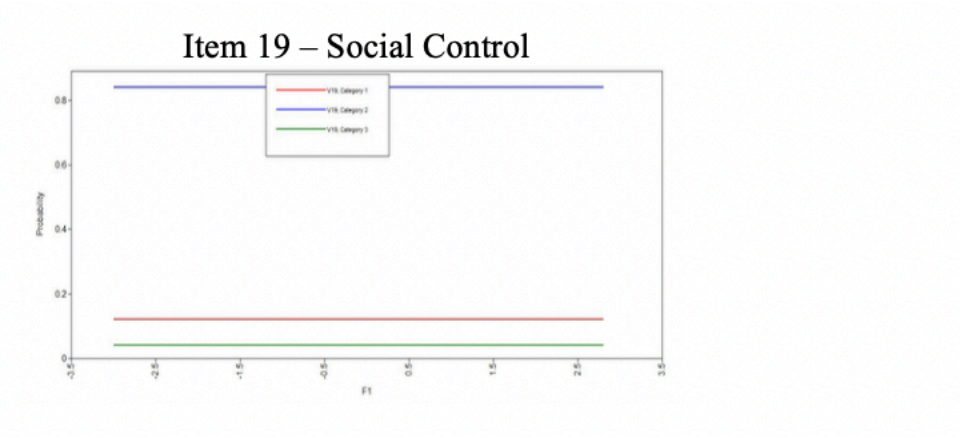


Figure 7. Continued.



Note. CRCs can be interpreted as the probability that a JIP with a given trait level (i.e., how applicable the item actually is to them [measured in deviations from the mean on the X axis]) will actually be given a score corresponding to that level of need (categories 1, 2, and 3 represent scores of 0, 1, and 2).

Figure 7. Continued.

Examination of the CRCs in Figure 7 coupled with the IICs above suggests that five DRAOR items (items 2, 5, 6, 8, and 9) account for much of the information and discriminative ability provided by the scale. These findings have a number of important implications, especially with respect to informing the refinement of the DRAOR in future studies. In the context of the present project, the purpose of these analyses was to determine if there was a small number of problematic items that might be contributing to the findings of measurement non-invariance. Evidently, there are more than a few potentially problematic items and removing all but five prior to re-running invariance testing analyses was not practicable.

Discussion

The goal of Study 1 was to evaluate the psychometric properties and factor structure of the DRAOR in a sample of racially diverse JI women serving community supervision orders in Iowa. Inconsistencies in the factor structure of the DROAR have emerged in prior research (i.e., Chadwick, 2014; Hanby, 2013; Yesberg & Polaschek, 2015), and Study 1 provided a unique opportunity to investigate what structure might emerge in a sample of 2,591 White, Black, and Hispanic JI women. As described earlier, the DRAOR is divided into three subscales, based on two features: (a) whether items theorized to change more rapidly or more slowly (Stable vs. Acute risk factors); and (b) the direction of the association between the items and recidivism (protective factors vs. risk factors). The three projects mentioned above used similar factor analytic strategies but yielded different results. Chadwick's (2014) exploratory factor analysis (EFA) returned a two-factor structure, with all of the risk items loading together, separate from

protective factors. Hanby's (2013) analyses suggested that the original DRAOR structure provided adequate fit for her data, but that model fit could be improved by allowing some risk items to cross-load or by moving some of the items from the Stable subscale to the Acute subscale and vice-versa. Yesberg and Polaschek (2015) examined the DRAOR's original three subscale structure using a confirmatory factor analysis (CFA) approach and found that it did not fit the data from their high-risk sample well. Similar to Hanby (2013), their results suggested some crossover between the Stable and Acute risk subscales. Moreover, Yesberg and Polaschek (2015) found that the items included in the Acute risk subscale loaded conceptually onto two separate subscales defined by either internal features or external circumstances, resulting in a four-factor structure. Considered collectively, findings from prior DRAOR research support the existence of a cohesive Protective construct but cast doubt on the theorized distinction between Stable and Acute risk factors.

Ideal Factor Structure

The first step in Study 1 involved psychometric and exploratory analyses which were conducted on subset (n = 500) of women randomly selected from the larger sample. While the three original DRAOR subscales demonstrated adequate internal consistency assessed via Cronbach's α, the inter-item correlations for the Acute and Stable subscales were low. By contrast, the inter-item correlations for the Protective subscale were consistently strong. Conspicuously, the results of these preliminary psychometric explorations echoed the results of the DRAOR studies described above and presaged subsequent findings regarding the DRAOR's factor structure in this study. Results of the EFA did not support the original three-factor structure and instead suggested two

alternative models – another three-factor structure with two different risk subscales and four-factor structure with three risk subscales. Notably, the Protective subscale, emerged in its hypothesized (original) structure in both alternative models.

The next step involved selecting the best model to use for subsequent invariance testing. To do this, confirmatory factor analyses (CFA) and Exploratory Structural Equation Modeling (ESEM) approaches were applied to the two alternative models that emerged following EFA. For the sake of completeness, the fit of the original DRAOR structure was also assessed using CFA and ESEM models. Results of the CFAs indicated that the alternative three- and four-factor models provided a better fit for the data than original DRAOR three subscale structure. Both of these models provided excellent fit according to all assessed model fit indices, with very little difference between the three- and four-factor models. Because the factor structure was ambiguous following EFA and cross-loadings have been identified in prior research, three- and four-factor ESEM models were also estimated. Both ESEM models resulted in slightly improved objective fit over the CFA models but given that the CFA models already provided excellent fit for the data, these improvements were of negligible practical significance. Also, the ESEM models were considerably more complex.

CFA is a theory-driven approach in which individual items can only be related to one latent construct, and loadings onto all other factors are constrained to zero (Tsigilis et al., 2018). By contrast, ESEM allows cross-loadings and can estimate structural and measurement models simultaneously, effectively combining the functions of EFA and CFA. ESEM models are inherently less parsimonious and therefore more difficult to interpret, but in cases of poor CFA fit, finding an accurate measurement model should be

prioritized over finding a pure set of measurements (Asparouhov & Muthen, 2009). At this stage, the task was to select the model that offered the best balance of interpretability and model fit. In view of this, the three-factor CFA model was chosen. This model fit the data well and was the most parsimonious. The four-factor CFA model, though also a good fit, was less conceptually appealing, particularly given that one of the risk subscales only had two indicators. Moreover, the two items in question, interpersonal relationships and living situation are not conceptually related.

Notably, statistics describing internal consistency (i.e., Cronbach's alpha) and factor analytic procedures were developed to assesses whether a given set of items measure a single underlying latent construct. Accordingly, sole reliance on factor analysis when evaluating the utility of a risk assessment in problematic. Some (e.g., Baird, 2009) have criticized the appropriateness of these statistical procedures for application to risk assessment tools, noting that the primary purpose of a risk assessments, to predict reoffending, may be at odds with goal of factor analysis, to creating a measure that captures a conceptually coherent, unitary construct. This argument has merit; to maximize prediction, a more comprehensive, and therefore less unitary approach is typically required. As noted by Kroner et al., (2007), aggregating relevant variables into a comprehensive, multifaceted tool is most likely to maximize prediction. Additionally, leading theories of criminal behaviour (e.g., Andrews and Bonta's (2010) personal, interpersonal and community reinforcement perspective [PIC-R], Farrington's (2011) integrated cognitive antisocial potential theory [ICAP], and Simons and Burt's (2011) social schematic theory [SST]; see Lloyd (2015) for a description of these theories) conceptualize risk as a dynamic interplay between multiple internal features and external

circumstances, not as a single, cohesive construct. However, if an assessment tool purports to assess a specific construct or set of constructs, it should exhibit a corresponding factor structure.

In the case of the DRAOR, relying on factor analysis is arguably a limited approach. DRAOR items were chosen based on their demonstrated relationship with recidivism and/or their importance to case management as reflected in the transition model of offender change. Items are divided into subscales based on their hypothesized temporal relationship with recidivism and whether they are expected to increase or decrease likelihood of reoffending. Otherwise, there is no reason, conceptual or otherwise, that the items would necessarily fit together onto the three subscales comprising the original DRAOR. The DRAOR does not specifically purport to capture psychologically distinct constructs – instead, it aims to inform case management decisions. Nevertheless, the series of factor analyses described above justified moving forward with subsequent analyses using the alternative three-factor DRAOR structure while being cognizant that there may not be a clear conceptual distinction between the indicators comprising the two risk factors.

Measurement Invariance

Importantly, for this dissertation, it was less important that the DRAOR demonstrate a specific structure than a consistent structure. Measurement invariance, or factorial invariance across measurement occasions, refers to empirical evidence that scale items assess the same construct(s) across different measurement occasions, be it across different population subgroups, or different assessment times (e.g., baseline assessments and later re-assessments). Invariance across racial groups and over time (i.e., baseline and

last assessment) in the main sample were assessed using multi-group confirmatory factor analysis (MGCFA). The hypothesis that the DRAOR would meet the criteria for strong factorial invariance across racial groups was not supported. Both the original DRAOR structure and the best alternative model achieved only configural, or structural invariance. These findings impose severe limitations on how DRAOR scores can be interpreted. More precisely, these findings imply that comparisons across racial groups cannot be made; a given DRAOR score cannot be assumed to mean the same things for women from different racial backgrounds. Likewise, measurement invariance over time was not sufficiently supported to allow for confident comparisons between baseline and last assessment. Weak invariance was achieved for baseline and last assessment, indicating that variances and covariances could be meaningfully compared across time points; means however, cannot be compared.[12]

The lack of the measurement invariance between baseline and last assessment was somewhat of a surprise. Notwithstanding the inconsistencies found in the factor structure itself, Hanby's (2013) and Yesberg and Polaschek's (2015) findings were supportive of measurement invariance over time. Hanby (2013) conducted CFAs on baseline scores as well as reassessment scores at two time points using the original DRAOR structure. Results indicated that the original three-subscale structure remained a good fit at both reassessment time points. Similarly, Yesberg and Polaschek (2014) conducted three separate CFAs and reported that results did not differ significantly when analyzing either

[12] Measurement invariance analyses were also performed using the original the subscale structure of the DRAOR for comparison purposes. Invariance across groups and over time (i.e., baseline and last assessment) were evaluated. Results for the original three subscale DRAOR structure were likewise non-invariant, and model fit indices were poorer.

first or third DRAOR assessment. Their findings broadly support the assumption that DRAOR scores can be interpreted in the same way at first and later reassessments. However, there are more rigorous ways to test the assumption of measurement invariance, including the approach used in this dissertation. In any case, the more successful results of these studies raise questions about whether use of a more rigorous approach combined with an unusual sample (i.e., racially diverse JI American women) could have contributed to the lack of measurement found in this dissertation. Other possibilities, including potential differences in DRAOR implementation and staff variables, are discussed in a later section.

Exploring Non-Invariance

Given the importance of measurement invariance for making group comparisons, supplementary analyses were undertaken in attempt to identify possible causes. Category response curves (CRCs) and item information curves (IICs) were generated to explore how individual items contributed to scale functioning. The goal of these analyses was to identify problematic items and explore whether it might be possible to remove said items. Had this been the case, initial factor analytic analyses could have been re-run and measurement invariance would have been explored with the new structure. Results clearly demonstrated that this was not feasible. Only five of the 19 DRAOR items contributed significant information to the overall explanatory power of the scale, and the same five items were the only items to demonstrate discriminative power. In other words, the only items that meaningfully differentiated between scores of 0, 1, and 2 were attitudes toward authority, sense of entitlement, attachment with others, anger/hostility, and access to victims. As all DRAOR items were selected for inclusion in the instrument

based on an expected utility in case management and the contribution of individual items has yet to be formally examined, it is unclear why these five items would have emerged together as the most informative and discriminative items.

Interestingly, attitudes toward authority, sense of entitlement, attachment with others, anger/hostility are all internal factors, or characteristics that are present at the level of the individual. Access to victims, on the other hand, also has an external component. As suggested by the opportunity theory of crime (see Cloward & Ohlin, 2013), the occurrence of a crime depends on (a) the existence of a motivated offender (or offenders) willing to engage in a given criminal behaviour, and (b) appropriate environmental conditions (i.e., the opportunity). However, as noted by several researchers in the field (e.g., Cohen & Felson, 1979; Cohen, Felson, & Land, 1980; Hindelang, Gottfredson, & Garofalo, 1978, etc.), variations in individuals' life-style and/or routine activities can significantly affect opportunities for crime. Thus, this item may still be strongly impacted by individuals' internal characteristics.

While it would be possible to generate potential explanations for why these items and not others provided more information and better discrimination with the current sample, attempts would be largely speculative. Increasingly, results of the current study diverge from those of prior research, raising the question of whether the instrument or characteristics of this sample are the explanation. Given that this was the first of three studies, and that Studies 2 and 3 aim to explore different aspects of the DRAOR's predictive ability, conclusions will be discussed later.

Implications for Practice

The single most significant implication for Study 1 was that, contrary to expectations, the DRAOR did not demonstrate adequate measurement invariance across racial groups or time. As only configural invariance was achieved, DRAOR scores cannot be confidently compared across JI women with different racial backgrounds or over time (i.e., comparing first and last assessment scores). Although it is still valuable to examine the predictive validity of the DRAOR subscales and items in this sample of women as this information is valuable for its own sake, measurement non-invariance precludes comparisons between the different racial groups in the sample, as DRAOR scores cannot be assumed to represent the same level of risk or protection across groups. For example, if both a White woman and a Black woman receive the same DRAOR score, we still cannot assume that they are equally likely to reoffend. Similarly, the same DRAOR score at first and last assessment cannot be assumed to indicate the same likelihood of recidivism, even for JIPs of the same race. Accordingly, comparisons across groups cannot be made in Studies 2 and 3, and instead, analyses will apply only to the specific race and assessment that they are conducted on. While findings will still be informative, subsequent studies will not be able to make general recommendations for dealing with all women being supervised in the community. The predictive validity of the DRAOR (Total, subscale, and individual item scores) may vary as a function of racial background and assessment time; case managers will need to consider this information carefully.

Limitations and Future Directions

The results of this study should be considered in light of two main limitations. First, there was significantly variability in the number of women in the three different

racial groups examined. Though it is not unusual to have a much larger proportion of White JIPs in an Iowa correctional sample, given that a goal of the current research was to explore difference across racial groups, having a larger number of individuals in each of the historically marginalized groups examined would have improved the reliability of results. The overall sample size in Study 1 was quite large (n = 2,768) and the sample size of the White and Black groups (n = 2,418 and 560, respectively) were sufficiently large, there were only 113 Hispanic women. According to sample size guidelines put forth by Comrey & Lee (1992), sample sizes of 100 are considered poor. Factor analytic techniques are large-sample techniques and correlation coefficients tend to be less reliable when estimated from small samples (Tabachnick & Fidell, 2013). Accordingly, larger samples of Hispanic women and women belonging to any other historically marginalized groups of interest would strengthen analyses.

Second, with no information available about the parole officers it is virtually impossible to make any conclusions about reasons why these findings might occur. Aside from underlying differences in this sample (the internal characteristics and external circumstances might differ substantially from samples used in the past) there are a number of staff variables that could also affect the data. For example, variables such as parole officers' gender and perceptions of JI women could be expected to have an impact on how the DRAOR was scored. Seng and Lurigio (2005) investigated probation officers' views on supervising women probationers and found that, officers (both men and women) reported that female clients tended to be more challenging to supervise than men. Specifically, the surveyed probation officers reported that female probationers tended to present with more problems and obstacles (e.g., childcare, trauma, financial,

personal/mental health problems, etc.) and required more time and energy on account of these issues. The greater emotionality of female probationers was also cited as a challenge. According to Seng & Lurigio (2005), male probation officers often reported difficulties communicating with, and thus changing, female clients' attitudes and behaviours. Feelings of frustration when dealing with female clients' "attitudes" (e.g., resistant, confrontational, passive-aggressive, manipulative, non-responsive) were also common. Conversely, female probation officers reported few difficulties with "attitude" and were more comfortable listening to extensive histories of personal problems. Female probation officers' responses suggested that they had fewer issues dealing with women's emotionality and that the major difficulty they encountered was insufficient time and resources to adequately respond to the needs of their female clients. Information regarding the gender of the parole officers in the current study as well as some understanding of their perceptions of women probationers and parolees may have helped clarify findings.

Other potential fidelity issues include quality of staff training and individual comfort scoring the tool (Desmarais and Singh, 2013), as well as the extent of "buy-in" (Latessa & Lovins, 2010; Miller & Maloney, 2013; Schlagger, 2009). For example, Miller & Maloney (2013) examined individual officers' reasons for compliance and found that that those who completed the tool because they believed the tool was useful were more accurate when compared to other officers who completed the tool to meet formal requirements. Similarly, in their examination of the STABLE-2007 and ACUTE-

2007, Hanson and colleagues (2007) found a high level of predictive accuracy for the assessments conducted by conscientious officers. [13]

Finally, previous studies (e.g., Adelman, 2020; Matz et al., 2018) have noted that parole officers are often pressed for time, which could certainly affect scoring fidelity. Some DRAOR items (e.g., problem solving skills, responsiveness to advice) require substantial background information and understanding of an individual's circumstances which may be difficult to collect in a short time, especially if the quality of the working relationship between the supervision officer and releasee is lacking.

Beside simply continuing to research the impact that gender and race have on community supervision, future research should aim to explore how staff variables may affect data quality. For example, in the context of the current study, it would be interesting to explore whether there are any regional differences in scoring. It is reasonable to expect a certain amount of variability in scoring across individuals, and larger, jurisdictional difference might also emerge. With more than 1,100 staff members responsible for more than 30,000 JIPs supervised in the community across eight different judicial districts in fiscal year 2019 (Iowa Department of Corrections, 2019), exploring such differences could be very informative. As noted by Schaefer and Williamson (2018) and Viljoen et al., (2019) factors such as level of experience, training, and belief in the

---

[13] 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. Ottawa: Public Safety Canada.
Hanson et al., 2007 defined "conscientious" as those who submitted all of the requested information. Officers in the study were part of the Dynamic Supervision Project, a large, Canada-wide test of risk assessment methodologies with a sample of sex offenders. Participating officers completed training and then submitted the data they collected during their routine supervision meetings with sex offenders, sometimes for a period of up to three years.

utility of the DRAOR that could have impacted rating fidelity and could also have influenced how they responded to observed changes in JIPs on their caseload. Accordingly, examinations of (a) the quality and availability of DRAOR scoring training; (b) inter-rater reliability; (c) staff "buy-in" and compliance monitoring; (d) the impact of case manager – client relationships; and (e) how DRAOR scores are used to inform intervention strategies is recommended.

Conclusion

Study 1 explored the factor structure and extent of measurement invariance of the DRAOR in a sample of JI women serving community supervision orders in Iowa. The factor structure that provided the best fit to the data contained three subscales, but the two risk subscales represented a different combination of items from the original DRAOR structure. Notably, however, the Protective subscale emerged in its original form and demonstrates conceptual and statistical cohesiveness. Despite fitting the data well, the new DRAOR structure did not demonstrate adequate measurement invariance across groups or over time. Thus, DRAOR scores cannot be meaningfully compared across racial groups or assessments.

Of the 19 DRAOR items, five were found to account for the vast majority of the discriminative ability and predictive information provided by DRAOR Total scores. If taken at face value, the performance of the items would suggest a need to go "back to the drawing board" so to speak. Considered more carefully, these findings raise questions about the fidelity of the data; the DRAOR items were derived from the transition model for improved case management, and as such, the expectation was that they would have a demonstrable relationship with recidivism. At this juncture, it is not possible to speculate

on whether intersectionality is involved; a later study that involves a matched sample of men (Study 3) and more specifically examines the DRAOR's predictive ability at both the domain- and item-level should shed some light on this issue.

Chapter 6: Study 2 – Examining the Sample

Purpose

Racial and ethnic historically marginalized groups are overrepresented in prisons in North America (Carson, 2015), and this is especially problematic for Black JIPs in Iowa (IDOC, 2017). When attempting to understand the causes of this disproportionality, it is important to acknowledge that disparities in per capita incarceration rates are likely to reflect the intersection of several different factors, namely: (a) racial differences in the prevalence, incidence, and nature of criminal behaviour; (b) actual bias or discrimination; and (c) the unequal influence of apparently race-neutral policies (e.g., in arrests, sentencing, case-processing, etc.) throughout the criminal justice system (Frase, 2013).

Study 2 explored issues related to the first of these three factors by investigating whether (a) post-release outcomes (i.e., violations, new offences, and any return to custody) and (b) time to recidivism (i.e., survival time) varied significantly across White, Black, and Hispanic JI women. One of the criticisms of mainstream correctional practice is that it often fails to account for variation in the correctional experiences of individuals belonging to historically marginalized groups. Significant differences in the base rates and survival times of different race groups would have important implications for case management. The following two hypotheses were made:

Hypothesis 1. Recidivism rates will be highest for Black women. Early scholars (e.g., Adler, 1975; Pollack, 1950; Smith & Visher, 1980; Sutherland & Cressey, 1978) speculated that Black women and White men committed crimes at a similar rate. Cooper and Smith (2011) investigated this hypothesis (the Black women/White men crime convergence hypothesis) and found little support; although Black women had higher rates of homicide when compared to White women, this rate was nowhere near that of White men. Notably, this study only examined homicide rates and was focused on offending rather than re-offending. Nevertheless, these findings, combined with the overrepresentation of Black individuals in the American prison population (Carson, 2015) suggest that rates of technical violations and new offences could be expected to be higher for Black women when compared to White and Hispanic women[14]. Base rates were expected to be similar for White and Hispanic women.

Hypothesis 2. Similarly, differences in survival times were expected for Black women relative to White and Hispanic women when controlling for static and dynamic risk. No differences in survival time were expected between White and Hispanic women. Survival time was also expected to vary as a function of risk. Specifically, women with more intensive levels of supervision (assigned based on IVVI scores) and higher dynamic risk (as assessed by the DRAOR) were expected to fail significantly more quickly than their lower risk counterparts.

[14] The current study does not propose to examine this hypothesis as there is no male comparison group. Interest in this hypothesis is limited to the implications that base rates might be expected to be higher for Black women.

Method

Participants

The sample utilized in Study 2 was comprised of 3,091 White (*n* = 2,418), Black (*n* = 560), and Hispanic (*n* = 113) JI women on community supervision orders in Iowa. For complete demographic information (i.e., age, marital status, education), level of supervision (i.e., static risk) and dynamic need, see the sample description provided in Chapter 4.

Procedure, Measures, and Outcome Data

The overall procedure, relevant measures (DRAOR, IVVI), and outcome variables of interest (technical violations, new offences, any return) are also described in Chapter 4.

Analytic Approach

Analyses for Study 2 were conducted in SPSS (Version 25) and Microsoft Excel. To increase interpretability of results, categorical risk bins rather than continuous DRAOR Total scores were used. The risk bins in question were developed by Perley-Robertson, Chadwick, and Serin (2020) and correspond to cut-points in DRAOR Total scores that discriminated groups of JIPs with significantly different recidivism rates. The risk bins and associated range of scores are: low-moderate (-12 to 2), moderate (3 to 9), moderate-high (10 to 22), and high (23+).[15] Level of supervision (LoS), the other covariate, was already categorical, thus, no recoding was necessary. Assumptions specific

[15] The DRAOR was not developed for use with low or very low offenders; hence, there is no 'low' group (Serin, 2007).

to Cox regression survival analysis (i.e., proportionality of hazards) were examined prior to analyses.

Calculating base rates. Base rates were established for violations, new offences, and any returns for each of the subgroups of interest (women identifying as White, Black, or Hispanic). Chi-square tests of independence were used to determine if base rates were significantly different across the three race groups. Briefly, the Chi-Square statistic (χ^2) is a non-parametric (distribution free) tool designed to assess group differences when both the independent and dependent variables are measured at a nominal level. Thus, it is ideal for comparing recidivism rates, measured dichotomously, across JIP subgroups, measured nominally. The Chi-square is a significance statistic, and should be accompanied by a strength statistic, such as the Cramer's *V* (McHugh, 2013). The Cramer's *V* is a form of correlation and is interpreted in the same way. Notably, when the outcome is only partially dependent on the independent variable, the strength correlation obtained is expected to be weak (McHugh, 2013). This is important to keep in mind for the current study as race is not expected to have a causal role in recidivism.

Survival analysis. Survival analyses are statistical procedures for measuring the amount of time until an event occurs (i.e., the amount of time between release and recidivism), and Cox regression is a function of survival analysis that applies regression methodology to survival data (Tabachnick & Fidell, 2013). Cox regression survival analysis (see Singer & Willet, 2003; Willet & Singer, 1991, 1993; Willet, Singer, & Martin, 1998) is a widely used model in correctional psychology and is currently viewed as the gold-standard for modeling the relationship between predictor variables and later dichotomous outcomes. Cox regression survival analysis also allows multiple predictors

to be entered simultaneously into a model to determine their unique and independent contributions to an outcome variable (Tabachnick & Fidell, 2013). This approach was selected over the Kaplan-Meier method for several reasons. First, Cox regression readily incorporates variable follow-up times and sample censoring (i.e., when there is incomplete information about an individual's survival time; Brown et al., 2009) while also taking additional explanatory variables (covariates) into account (Bewick, Cheek, & Ball, 2004). Second, Kaplan-Meier survival analysis is a parametric procedure and requires a single binary predictor; Cox regression, on the other hand, is a semi-parametric procedure and readily incorporates both binary and continuous predictors (Bewisk et al., 2004). As per Lloyd (2015), Cox regression survival analysis is uniquely appropriate for predicting events that terminate participation in a study (i.e., recidivism leading to reincarceration) while still preserving all of the data collected on these participants prior to termination of participation *and* retaining information collected for individuals who did not experience the predicted outcome (i.e., remained crime free and are right-censored). Further, as examining the impact of risk on survival time was of interest for this study, Cox regression was determined to be the best approach.

Survival analysis controls for time-at-risk by including it in the outcome of the test, whereby both the time to an event and the proportion of a group experiencing an event are considered in the hazard of an event. Hazard ratios can be interpreted as the change in hazard with a single unit of change in the associated variable. A hazard ratio of 1 indicates no change in the hazard of an event, whereas hazard ratios greater than 1 indicate that the hazard of experiencing an event increases as a variable increases or that the hazard in one group is greater than the other group. Hazard ratios less than 1 indicate

that the hazard of an event decreases as a variable increases or that the hazard of one group is less than that of the other group.

The reporting of effect sizes alongside results of regression analyses is important with respect to ease of interpretation and generalizability of findings. As noted by Lloyd (2015), researchers utilizing Cox regression models typically only report odds ratio (*OR*) values generated by the model to describe the strength of a finding, which creates problems when attempting to generalize across studies using different measures. Moreover, the most commonly used statistics packages have not of yet incorporated effect size calculation options for running Cox regression analyses, and consequently, effect size options are somewhat limited. The R^2 measure of association utilized in this study was originally provided by Cox and Snell (1989) for logistic regression and was demonstrated for survival analysis by Allison (1995). This measure is based on G^2, which is a likelihood-ratio chi-square statistic (Tabachnick & Fidell, 2013). Models are estimated both with and without covariates, and a difference (G^2) is calculated based on SPSS Cox regression output using the following equation:

$$G^2 = [(-2 \text{ log-likelihood for smaller model}) - (-2 \text{ log-likelihood for larger model})]$$

Next, the R^2 value is calculated as

$$R^2 = 1 - e^{(-G2/n)}$$

Based on Allison (1995), this R^2 is not the proportion of variance in survival that is explained by the covariates, but instead represents the relative association between survival and the covariates tested.

Results

Data Management and Assumptions Tests

Important predictor variables in this analysis were race (White, Black, and Hispanic), IVVI level of supervision (LoS) ratings (a proxy for static risk level), and baseline DRAOR scores (a measure capturing dynamic risk and need). To examine survival time, three variables measuring time to failure for each outcome were computed. Time to violation, new offence, and any return were calculated using each JIPs' release date as a starting point and were calculated in months to failure. Successful cases (i.e., cases where no community failure occurred before the end of data collection) were right-censored, meaning the study end date (01-May-2019) was used. The existence of a considerable number of these cases was anticipated prior to study commencement and was an important factor underlying the decision to select cox regression as an approach in the first place – it seamlessly incorporates this information instead of having to truncate the follow-up and deal with the loss in sample size. In total, 1,017 cases (32.9%) cases were right censored for the technical violations outcome, 2,762 (89.4%) for the new offence outcome, and 965 (31.2%) for the any return outcome.

Basic assumptions were examined for baseline DRAOR Total scores as part of Study 1and were therefore not repeated here. As a brief summary, no concerns were identified. None of the covariates had missing data, and overall, sample size is sufficient. Assumptions regarding linearity, normality, and multicollinearity were met and no problematic outliers (univariate or multivariate) were identified. Accordingly, the final assumption tested for this study was the proportionality of hazards assumption specific to Cox regression survival analysis. When Cox regression is used to analyse differences

between levels of a discrete covariate (e.g., race), it is assumed that the shapes of the survival functions will be the same for each group over time (Tabachnick & Fidell, 2013). When this assumption is met, the survival functions for each group are roughly parallel and violations indicate a significant interaction between a given covariate and time.

Several approaches were necessary to comprehensively test the proportionality assumption. Examination of log minus log plots of the hazard functions is often used as a starting point when assessing this assumption, but results can be misleading. Thus, inspection of additional graphs such as plots of Schoenfeld residuals and formal tests for interactions between the time measures (i.e., time to each outcome) and model covariates are also required (Tabachnick & Fidell, 2013). In the current study, examination of the log minus log plot of the hazard functions suggested that no interactions were present for the race and level of supervision covariates, but the hazard functions for the DRAOR Total score risk bins were not parallel, suggesting a violation of the proportionality assumption for this predictor (see Appendix E for the log minus log plots). However, consideration of the additional metrics (i.e., inspection of Schoenfeld residuals and statistical evaluation of interaction terms) suggested otherwise. Examination of the Schoenfeld residuals (also known as partial residuals) for DRAOR Total scores plotted against time revealed a random pattern (indicating that the assumption held) and inclusion of the interaction term (DRAOR Total scores by time; see Block 2 in each of the models shown in Table 12 below) did not significantly improve the omnibus models. If any of the covariates did in fact interact with time, the Chi-square values evaluating the change in model fit from the previous step would be expected to achieve significance,

which was not the case. Results of tests for covariate by time interactions were consistent (i.e., non-significant) for technical violations, new offences, and any returns.

Graphs plotting Schoenfeld residuals against time and interaction terms for each level of each of the three categorical covariates (e.g., low-moderate DRAOR scores X Time, moderate DRAOR scores X Time, moderate-high DRAOR scores X Time, etc.) were also evaluated for each of the three recidivism outcome variables (see Appendix F for the regression coefficients for these interactions). None of these interaction terms were significant, even before adjustments to *p* values were made to control for family-wise error. Considered collectively, results indicated that the assumption of proportional hazards was upheld and that it was not necessary to treat DRAOR scores as time-dependent in the model.

Table 12

Cox Proportional Hazards Model Assumptions for Race, Level of Supervision, and DRAOR Scores

			Overall			Change from Previous Step		
		-2 Log Likelihood	Chi-Square	df	Sig.	Chi-Square	df	Sig.
Technical Violation	Omnibus Test of Model Coefficients	31174.809	156.933	9	.000	166.812	9	.000**
	Block 2 – Interaction Terms Added	31167.083	163.439	18	.000	7.726	9	.562$^{n.s.}$
New Offence	Omnibus Test of Model Coefficients	5167.20	67.77	9	.000	75.06	9	.000**
	Block 2 – Interaction Terms Added	5157.88	76.70	18	.000	9.32	9	.409$^{n.s.}$
Any Return	Omnibus Test of Model Coefficients	31854.45	186.58	9	.000	197.56	9	.000**
	Block 2 – Interaction Terms Added	31846.09	193.85	18	.000	8.361	9	.498$^{n.s.}$

Note. All omnibus tests were significant, denoted ** at the $p < .001$ level. The addition the interaction terms (each covariate by time) in Block 2 did not significantly improve the fit of the model, as demonstrated by non-significant (n.s.) changes from the previous step.

Base Rates

The purpose of Study 2 was to test whether the rates of violations, offences, and any returns as well as time to failure (i.e., time to these recidivism events) differed for White, Black, and Hispanic JI women. Average length of follow-up following release was longest for White women (*M* = 24.32 months, *SD* = 30.83), followed by Hispanic women (*M* = 23.36 months, *SD* = 19.56). The average follow-up time for Black women (*M* = 20.96 months, *SD* = 21.29) was significantly shorter when compared to White women (t (1170.02) = 3.076, p = .002), but not Hispanic women (t (139.44) = 0.497, p = .62). Base rates for technical violations, new offences, and any returns are shown in Table 13 below. Two thirds (67.1%) of women in the sample incurred a technical violation and 10.64% committed a new offence while under supervision in the community.

Table 13

Base Rates for Technical Violations, New Offences, and Any Return by Race

	Technical Violation *n* with violation (%)		New Offence *n* with offence (%)		Any Return *n* with return (%)	
All women (*n* = 3,091)	2,074	(67.10)	329	(10.64)	2,126	(68.78)
White (*n* = 2,418)	1,591	(65.80)	265	(10.96)	1,632	(67.49)
Black (*n* = 560)	413	(73.75)	57	(10.18)	422	(75.36)
Hispanic (*n* = 113)	70	(61.95)	7	(6.19)	72	(63.72)

As illustrated in Table 14, Black women were significantly more likely than their White and Hispanic counterparts to incur a technical violation. Correspondingly, they also had the highest rate of any return. White and Hispanic women did not have significantly different rates of technical violations and returns to custody. Contrary to expectations, Black women did not also have more new offences while on supervision;

rates of reoffending did not vary significantly as a function of race. Cramer's *V*, which compares the strength of the association between two variables and can be interpreted in the same was as Pearson's correlation coefficient, is very low for all comparison. This this suggests that race was not a primary determinant (i.e., cause) of reoffending behaviour even for the technical violation and any return outcomes.

Table 14

Chi Square Tests for Group Differences in Base Rates by Race and Community Outcome

		Chi-square (χ^2)	*df*	*p* (two-tailed)	Cramer's *V*
Technical Violation	White vs. Black	13.06*	1	< .001	.066
	White vs. Hispanic	0.71	1	.40	.017
	Black vs. Hispanic	6.47*	1	.01	.098
New Offence	White vs. Black	0.29	1	.59	.010
	White vs. Hispanic	2.56	1	.11	.032
	Black vs. Hispanic	1.73	1	.19	.051
Any Return	White vs. Black	13.14*	1	< .001	.066
	White vs. Hispanic	0.50	1	.48	.014
	Black vs. Hispanic	5.88*	1	.015	.094

Note. df refers to degrees of freedom, *p* (two-tailed) refers to significance level using a two-tailed test. Significance, (denoted *) was evaluated in relation to adjusted alpha values to control for family-wise error.

Base rates were also calculated as a function of level of supervision (LoS) and dynamic risk (DRAOR scores) and results for technical violations are displayed in Tables 15 and 16 below.[16] Importantly, this level of disaggregation resulted in very low sample size for several groups (i.e., the administrative and minimum groups for LoS, and the

[16] Base rates were also calculated for the new offence and any return outcome and are presented in Appendix G. The low base rate for new offences resulted in many suppressed cells and the findings for the any return outcome mirrored those for technical violations.

high DRAOR group), thus, cautious interpretation of results is required. Note that this caveat also applies to all analyses involving Hispanic women. Nevertheless, when broken down by LoS, important racial differences emerge; while the base rates for White and Hispanic women increase monotonically as level of supervision intensifies, for Black women, this is not the case. Base rates are highest for Black women supervised at the high normal level rather than the intensive level as would be expected. Moreover, the base rate seen for the low normal group is the same as the rate observed in the intensive supervision group. As a result, the base rate for Black women in the low and high normal groups are significantly higher than those of their White and Hispanic counterparts.

Table 15

Base Rates for Technical Violations Disaggregated by Race by Level of Supervision (LoS)

			Administrative (n = 30)		Minimum (n = 56)		Low Normal (n = 898)		High Normal (n = 899)		Intensive (n = 1,208)	
	n with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate
All Women (n = 3,091)	2,074	67.1%	8	26.7%	27	48.2	497	55.3%	633	70.4%	909	75.2%
White (n = 2,418)	1,591	65.8%	†		24	46.2%	376	52.1%	487	69.3%	699	76.1%
Black (n = 560)	**413**	**73.8%**	†		†		106	**72.6%**	121	**77.1%**	181	**72.7%**
Hispanic (n = 113)	70	61.9%	†		†		15	48.4%	25	64.1%	29	70.7%

Note. Cells with *n* of 5 or less are supressed (†), **bold** denotes results unexpected results.

Table 16

Base Rates for Technical Violations – Race by Dynamic Risk Assessment for Offender Re-Entry (DRAOR) Score

			Low – Moderate (*n* = 754)		Moderate (*n* = 1,132)		Moderate – High (*n* = 1,171)		High (*n* = 34)	
	n with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate	*n* with violation	Base rate
All Women (*n* = 3,091)	2,074	67.1%	424	56.2%	774	68.4%	855	73.0%	21	61.8%
White (*n* = 2,418)	1,591	65.8%	324	53.5%	569	67.3%	684	72.6%	14	**56.0%**
Black (*n* = 560)	413	73.8%	86	**71.7%**	177	**73.8%**	143	**74.5%**	7	87.5%
Hispanic (*n* = 113)	70	61.9%	14	50.0%	28	59.6%	28	75.7%	†	

Note. Cells with *n* of 5 or less are supressed (†), **bold** denotes results unexpected result

Similar issues were apparent when base rates were examined by race and DRAOR Total score. For White women, base rates increased proportionally to risk for the low-moderate, moderate, and moderate-high risk groups, but then decreased dramatically in the high risk group to 56.0%. However, given the small sample size for the high risk group (n = 34 across the whole sample), firm conclusions cannot be made regarding this result. For Black women, the base rate observed for those in the high risk group was more in line with expectations (i.e., > 80.0%), though similar to the results for LoS described above, little variability in base rates was apparent for Black women across all other DRAOR risk groups.

These findings provide limited support for the hypothesis that base rates would increase in relation to supervision intensity and level of dynamic risk. This appears to true for Hispanic women but is less clear for White and Black women. Instead, these findings suggest that there is a need to more carefully examine how Black women are assigned to levels of supervision and to explore explanations for why the White women assessed as high risk on the DRAOR are (a) assessed as high risk in the first place, and (b) incur technical violations at such a low rate.

Cox Regression Survival Analysis

Cox regression survival analysis was used to examine the effect of race and other covariates (static and dynamic risk) on the survival time of JI women on community supervision. Survival time was evaluated in relation to technical violations, new offences, and any returns. To control for risk, sequential Cox regression models were fitted, in which LoS and DRAOR scores were entered into the model first (as a set), followed by the race variable. This approach permits a likelihood-ratio test of the effect of race, after

statistical adjustment for the risk covariates (Tabachnick & Fidell, 2013). As both covariates were measured categorically, likelihood of survival (odds ratios) were compared across risk groups.

Technical violations. When entered simultaneously, the set of three variables (static and dynamic risk and race) significantly predicted survival time, with χ^2 (9) = 166.81, $p < .001$ and was therefore preferable to a null model (i.e., a model without the covariates). However, the associated R^2 value, .053, was very small, indicating that the strength of the association between the set of predictors and survival time was very weak. After controlling for static and dynamic risk, the effect of race on time to violation reached significance (χ^2 (2) = 8.87, $p = .01$) though these findings should be considered carefully in light of the obtained strength of the association between race and survival time ($R^2 = .003$).

Next, odds ratios (likelihood of technical violation) were calculated for women at each level of the risk and race covariates; results are summarized in Table 17.[17] Differences in survival time by race were apparent (see Figure 8). When controlling for risk, survival time was shortest for Black women and longest for Hispanic women. Consistent with hypotheses, Black women were significantly more likely (16.7%, *OR* =

[17] An alpha value of .006 was used to adjust for inflated familywise error rate for all survival analyses. The variables in the tables below (Tables 17, 18, & 19) represent dummy variables created so that differences between levels of each categorical covariate could be evaluated. For each covariate, one level (or category) is specified as the reference group and therefore has no *B* coefficient associated with it. The sign (+ or -) associated with the *B* coefficient indicates whether the likelihood of failure is higher (positive values) or lower (negative values) for offenders in that group relative to the reference group. Odds ratios (*OR*) can be interpreted as the likelihood of failure for offenders in a given group, relative to the reference group.

1.167) to incur a technical violation than White women. Survival time was not significantly different for White and Hispanic women.

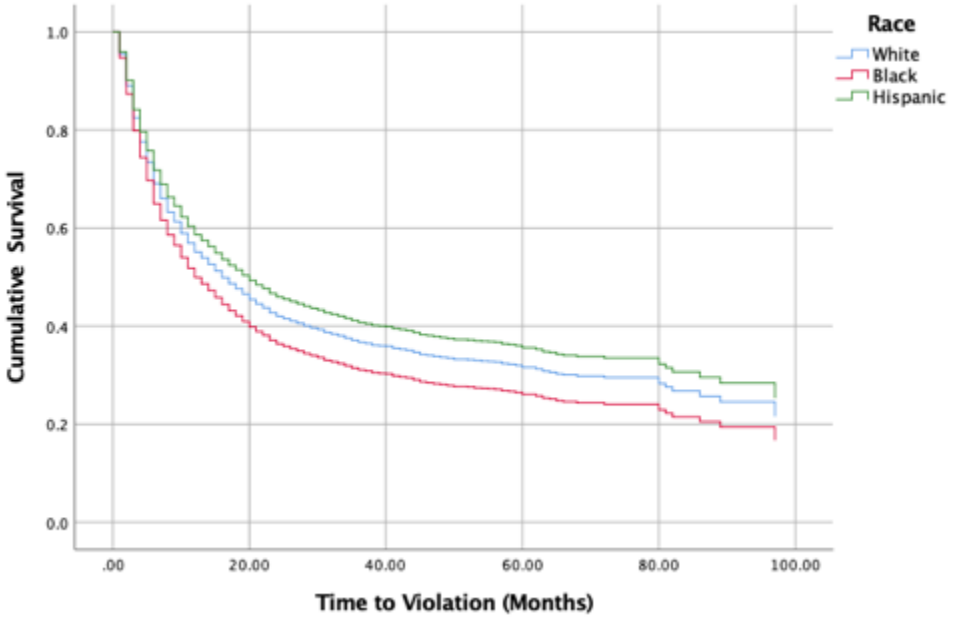


Figure 8. Time to technical Violation for JI Women by Race Controlling for Risk.

LoS. For LoS, odds ratios increased consistently as supervision intensity increased. Women supervised at high normal and intensive LoS were significantly more likely to commit a technical violation, with odds ratios of 3.44 and 3.90 respectively. In other words, women at these higher supervision levels were 3.4 (high normal LoS) and 3.9 (intensive LoS) times more likely fail than women supervised at the administrative level. Figure 9 illustrates the differences in failure risk according to LoS.

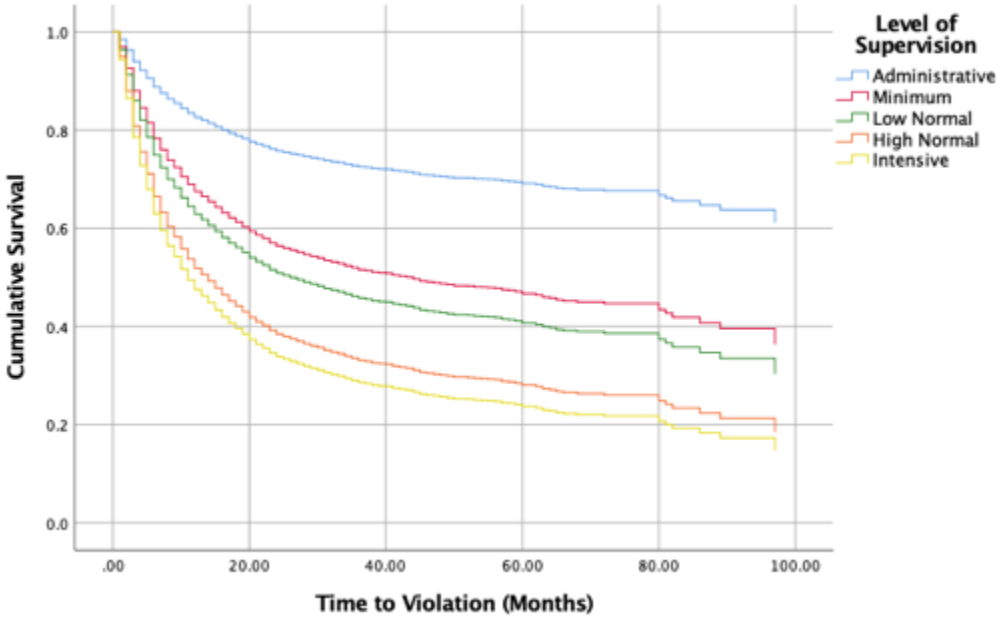


Figure 9. Time to Technical Violation for JI Women by Level of Supervision (LoS).

DRAOR scores. Findings regarding the effect of DRAOR scores on survival time were somewhat surprising. Specifically, women with highest DRAOR scores (i.e., the high risk group) were less likely to commit a technical violation than women in the reference group, (the low-moderate group; *OR* = 0.826; see Figure 10). Although the difference in survival time between the high risk group (*n* =21) and the low-moderate groups was not statistically significant and sample size for the high risk group was small, these findings nevertheless raise questions. Namely, why are women assessed as having the highest level of dynamic risk and need least likely to fail? This is unclear, especially as those in the second highest group (moderate-high) failed at 1.25 times the rate of those in the low-moderate group.

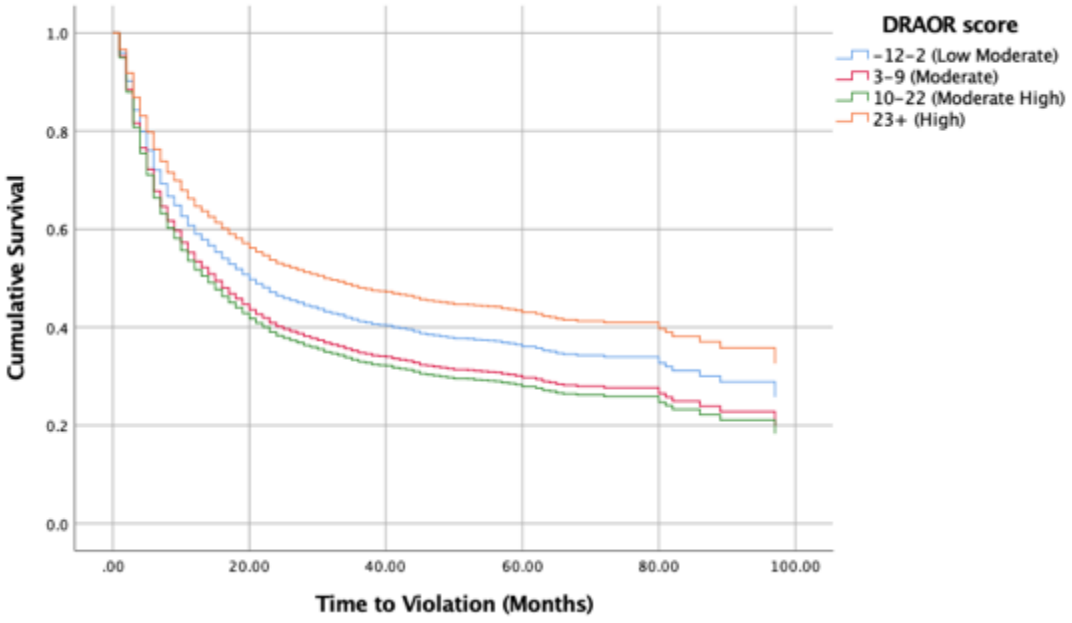


Figure 10. Time to Technical Violation for JI Women by DRAOR Score.

Table 17

Effect of Level of Supervision, Dynamic Risk Assessment for Offender Re-Entry Scores, and Race on Time to Technical Violation

	B	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI
LoS – Administrative [R]			69.733	4	.000[Sig]		
LoS – Minimum	.724	.403	3.232	1	.072	2.063	.937 – 4.541
LoS – Low Normal	.890	.357	6.219	1	.013	2.435	1.210 – 4.901
LoS – High Normal	1.237	.358	11.959	1	.001*	3.444	1.709 – 6.942
LoS – Intensive	1.362	.358	14.451	1	.000*	3.902	1.934 – 7.875
DRAOR – Low–Moderate [R]			14.039	3	.003[Sig]		
DRAOR – Moderate	.174	.065	7.167	1	.007	1.190	1.048 – 1.351
DRAOR – Moderate–High	.224	.068	10.902	1	.001*	1.251	1.095 – 1.429
DRAOR – High	-.191	.227	.708	1	.400	.826	.530 – 1.289
Race – White [R]			9.073	2	.011[Sig]		
Race – Black	.154	.056	7.708	1	.005*	1.167	1.046 – 1.301
Race – Hispanic	-.110	.122	.806	1	.369	.896	.705 – 1.139

Note. N = 3,091. [R] indicates the reference group, LoS = Level of Supervision, and DRAOR = Dynamic Risk Assessment for Offender Re-Entry (Serin, 2007) where *B* is the regression coefficient, S.E. is standard error, and *df* denotes the degrees of freedom for the analysis, and CI = confidence intervals. [Sig] indicates that the reference category of the covariate reliably predicts survival and * denotes significance for comparisons between the reference group and other levels of a covariate; alpha values of 0.006 was used to adjust for inflated familywise error rate.

New offences. The set of covariates (race, LoS, and DRAOR scores) improved prediction of time to new offence over the null model, with χ^2 (9) = 75.061, $p < .001$, though the strength of association between the covariates and survival time was too weak ($R^2 = .024$) to be impactful. Notably, there was no reliable effect of race on survival time after adjusting for the two risk covariates (χ^2 (2) = 3.463, $p = .177$). Thus, race contributes little to the prediction of new offences for the women in this sample. Results of the Cox regression survival analyses comparing time to new offence across levels of risk and race are presented in Table 18.

Figure 11 displays the survival curves for White, Black, and Hispanic women after controlling for static and dynamic risk. While survival time might appear to vary by race at first glance, careful inspection of the graph reveals that cumulative survival rates exceed 90% for all three racial groups. Contrary to hypotheses, survival time to new offence did not vary as a function of race and Black women did not have the highest odds of failure.

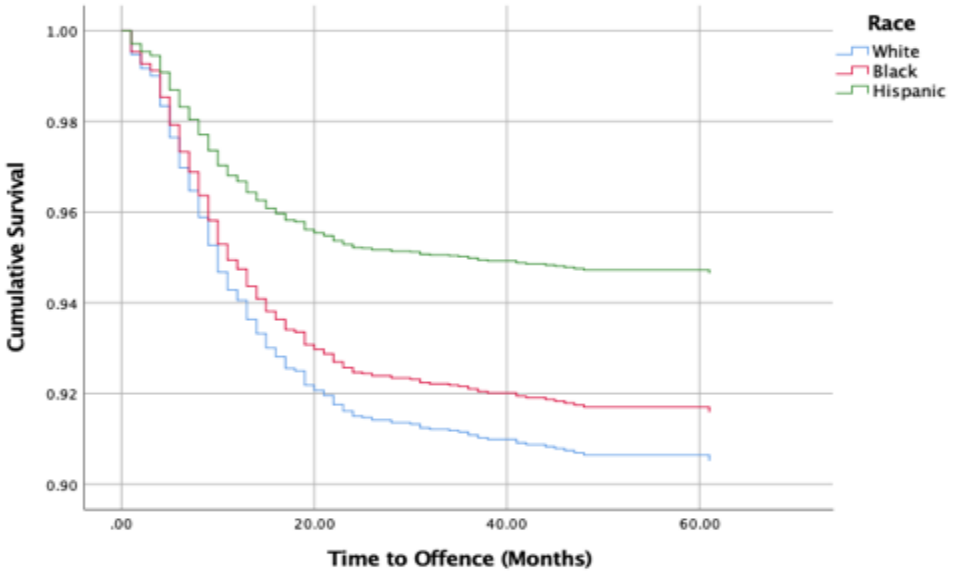


*Figure 11.*Time to New Offence for JI Women by Race Controlling for Risk.

LoS. Level of supervision was the only covariate that reliably predicted survival time for the new offence outcome (see Table 18). Notably, none of the women supervised at the administrative level and only one woman in the minimum group committed a new offence during the study period; to avoid artificially inflating odds ratios, the intensive supervision group was instead used as the reference category for group-level comparisons of time to failure. Women in the low, and high normal supervision groups did not differ meaningfully from those in the intensive supervision group, but women supervised at the minimum and administrative levels had significantly different (lower) rates of failure. Odds ratios of <.001 and .091 indicate that likelihood of committing a new offence decreased dramatically (i.e., by >99% and >90%) for women supervised at the minimum and administrative levels respectively.

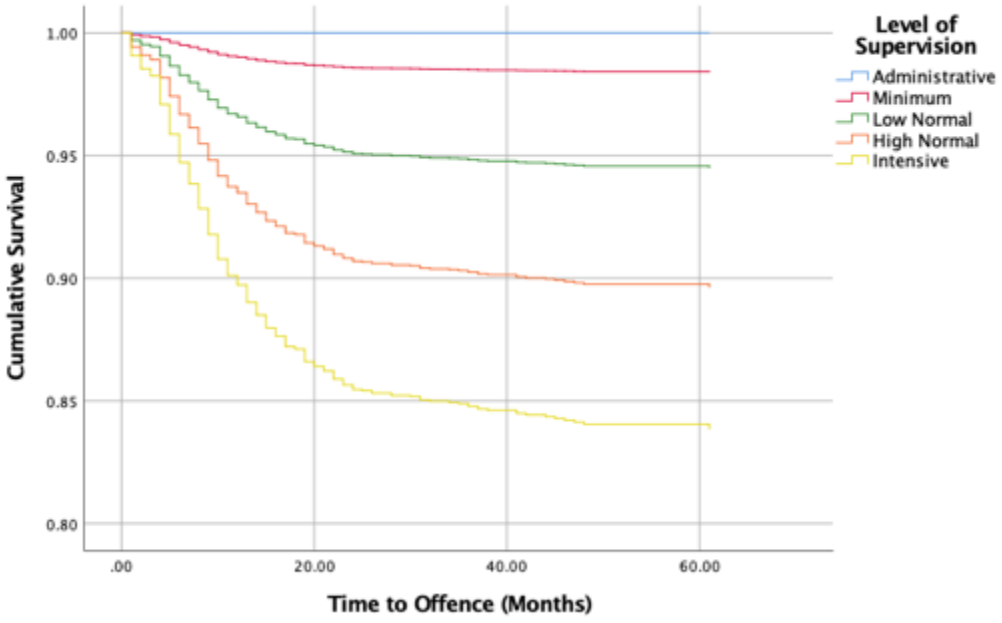


Figure 12. Time to New Offence for JI Women by Level of Supervision (LoS).

DRAOR scores. With respect to the impact of DRAOR scores on time to a new offence, results were inverse to expectations. Failure rates were highest for the lowest risk group and lowest for the group of women rated as having the highest level of dynamic risk and need (Figure 13). Though the differences in survival time across risk groups did not reach significance (see Table 18), the inversion of the anticipated survival curves for the low-moderate and high risk women was nevertheless surprising, notwithstanding the very small *n* for high risk women. The survival functions for the moderate and moderate-high groups were consistent with expectations. Considered in combination with the results of analyses predicting technical violations where the high risk group were also least likely to fail, it is clear that a closer examination of the case

management and supervision practices utilized with JI women at both ends of the dynamic risk continuum is necessary.

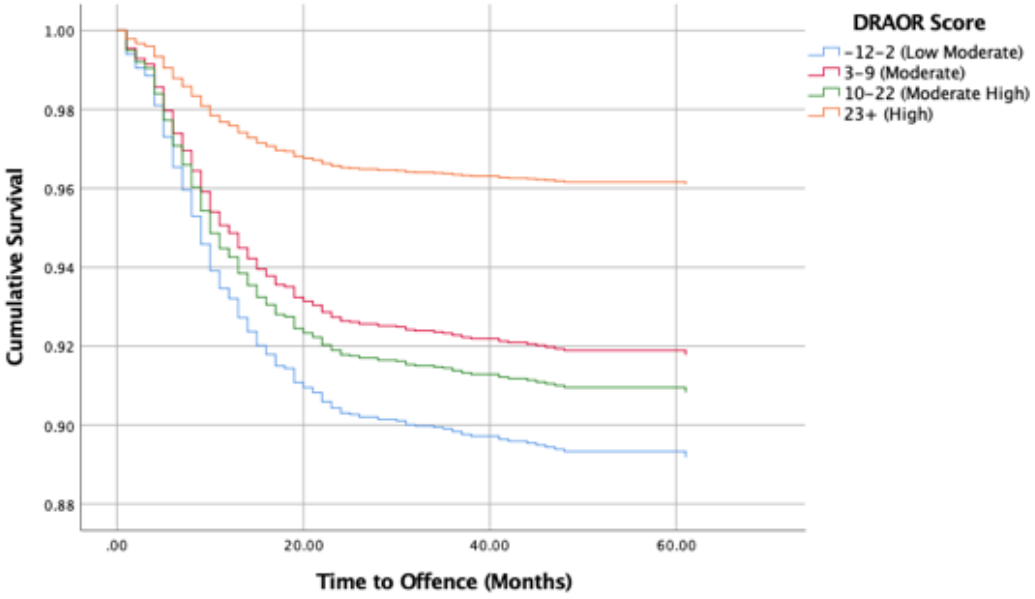


Figure 13. Time to New Offence for JI Women by DRAOR Score.

Table 18

Effect of Level of Supervision, Dynamic Risk Assessment for Offender Re-Entry Scores, and Race on Time to New Offence

	B	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI
LoS – Administrative	-.475	.134	12.568	1	< .001*	< .001	< .001 – 6.55E[67]
LoS – Minimum	-1.136	.175	42.038	1	< .001*	.091	.013 – .659
LoS – Low Normal	-2.393	1.008	5.632	1	.018	.321	.228 – .453
LoS – High Normal	-10.651	85.106	.016	1	.900	.622	.478 – .808
LoS – Intensive [R]			47.448	4	< .001[Sig]		
DRAOR – Low–Moderate [R]			4.550	3	.208		
DRAOR – Moderate	-.288	.166	2.996	1	.083	.750	.541 – 1.039
DRAOR – Moderate–High	-.174	.167	1.080	1	.299	.841	.606 – 1.166
DRAOR – High	-1.059	.723	2.143	1	.143	.347	.084 – 1.431
Race – White [R]			2.986	2	.225		
Race – Black	-.126	.146	.740	1	.390	.882	.662 – 1.175
Race – Hispanic	-.595	.383	2.415	1	.120	.551	.260 – 1.168

Note. N = 3,091. [R] indicates the reference group, LoS = Level of Supervision, and DRAOR = Dynamic Risk Assessment for Offender Re-Entry (Serin, 2007) where *B* is the regression coefficient, S.E. is standard error, and *df* denotes the degrees of freedom for the analysis, and CI = confidence intervals. [Sig] indicates that the reference category of the covariate reliably predicts survival and * denotes significance for comparisons between the reference group and other levels of a covariate; alpha values of 0.006 was used to adjust for inflated familywise error rate.

Any Return. As described earlier, the any return outcome captures technical violations and new offences. Given the disparity in base rates, this outcome is disproportionately affected by the relationship between technical violations and survival time. Notwithstanding, any return (or any reconviction) is an outcome frequently considered by correctional staff. Accordingly, it was deemed important to evaluate.

The set of risk and race covariates significantly predicted any return, with χ^2 (9) = 197.563, $p < .001$, $R^2 = .061$. After adjusting for risk, race had a significant effect on survival time (χ^2 (2) = 7.824, $p = .02$), though again, the strength of the association with survival was very weak ($R^2 = .003$). This indicates that while the inclusion of the race variable in the model improved prediction of survival time compared to a model without it, that there was little meaningful difference in survival time across race groups. As shown in Figure 14, the survival curves of White, Black, and Hispanic women after controlling for static and dynamic risk (LoS and DRAOR scores) were very similar. Statistical comparisons of survival time (see Table 19) indicated that no significant differences were found between White, Black, and Hispanic women.

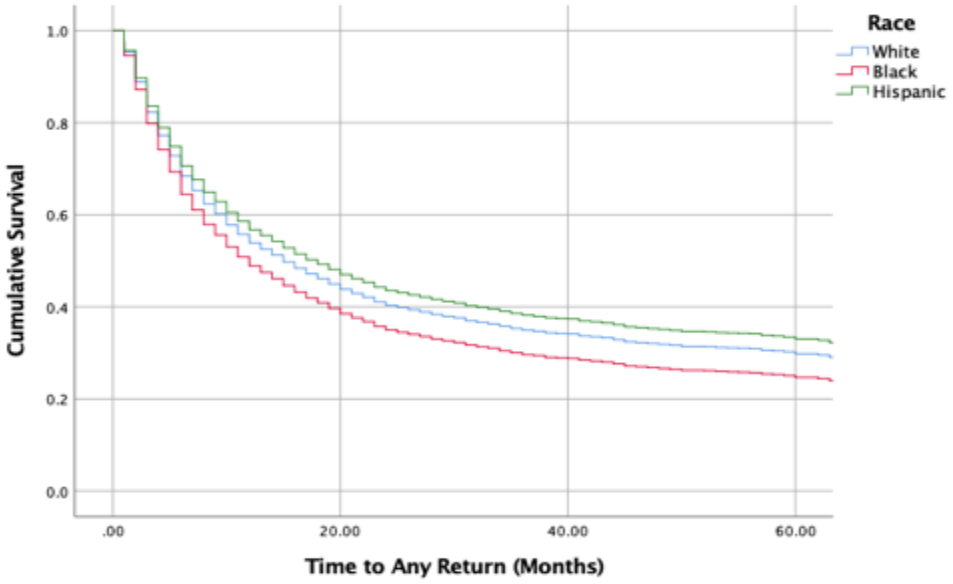


Figure 14. Time to Any Return for JI Women by Race Controlling for Risk.

LoS. Level of supervision had a reliable effect on survival time for the any return outcome. Compared to women supervised at the administrative level, those supervised as high normal and intensive were significantly more likely to fail, even after corrections were made to adjust for inflated familywise error (see Table 19). The difference in the survival curves for these groups is illustrated in Figure 15. Odds ratios of 3.63 and 4.25 indicate that those in the high normal and intensive supervision groups were significantly more likely to return to prison.

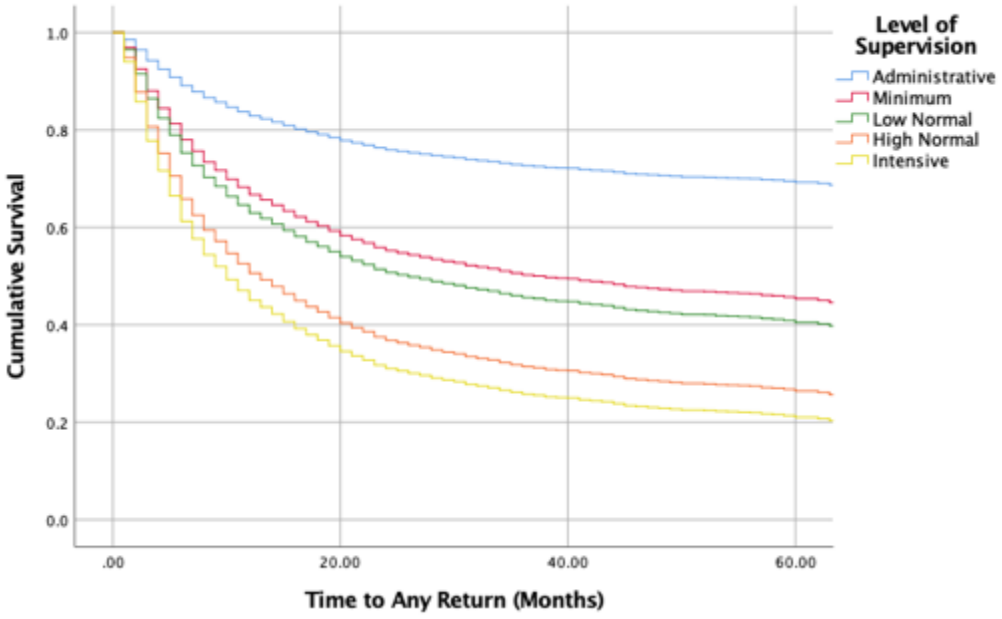


Figure 15. Time to Any Return for JI Women by Level of Supervision.

DRAOR scores. Overall, DRAOR scores also had a reliable effect on survival time for this outcome. Consistent with the pattern of survival curves for DRAOR risk categories seen for the technical violations outcome, women in the highest risk group had the highest survival rates though differences in survival time between the high risk group and the low-moderate group (the reference group) were not statistically significant (see Figure 16 and Table 19). Again, it should be noted that the sample size for the high risk group was small. The moderate-high risk group, however, which actually had the lowest survival rate, did differ significantly when compared to the low-moderate group, with an

odds ratio of 1.232.

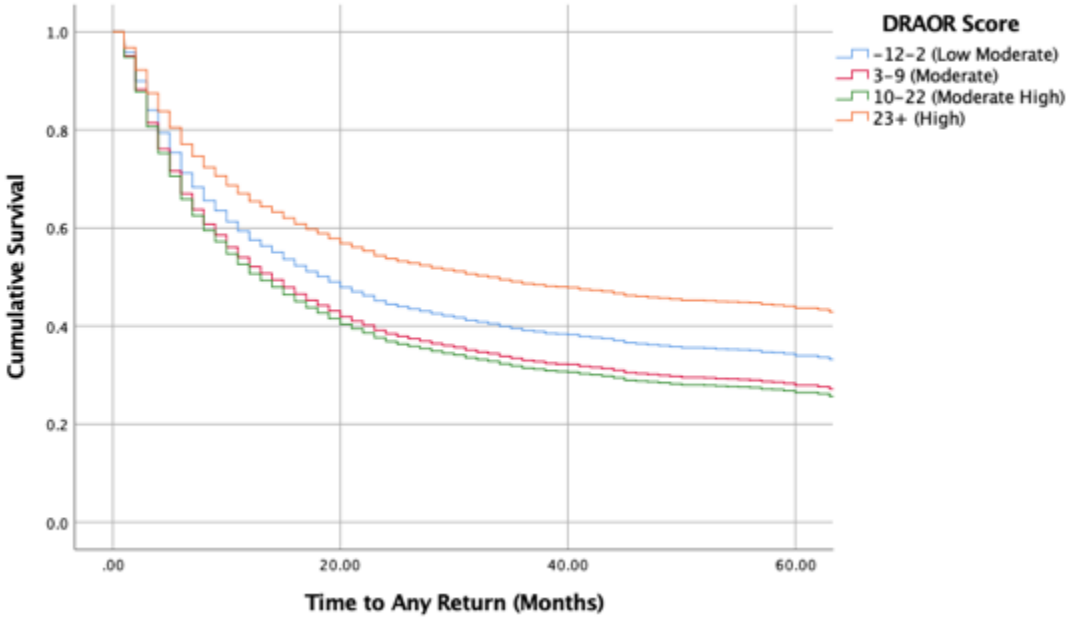


Figure 16. Time to Any Return for JI Women by DRAOR Score.

Table 19

Effect of Level of Supervision, Dynamic Risk Assessment for Offender Re-Entry Scores, and Race on Time to Any Return

	B	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI
LoS – Administrative [R]			90.156	4	< .001[Sig]		
LoS – Minimum	.767	.401	3.663	1	.056	2.154	.982 – 4.728
LoS – Low Normal	.900	.357	6.367	1	.012	2.461	1.223 – 4.952
LoS – High Normal	1.288	.358	12.985	1	< .001*	3.627	1.800 – 7.310
LoS – Intensive	1.446	.358	16.313	1	< .001*	4.247	2.105 – 8.569
DRAOR – Low–Moderate [R]			13.918	3	.003[Sig]		
DRAOR – Moderate	.166	.064	6.703	1	.010	1.181	1.041 – 1.339
DRAOR – Moderate–High	.209	.067	9.639	1	.002*	1.232	1.080 – 1.405
DRAOR – High	-.265	.227	1.370	1	.242	.767	.492 – 1.196
Race – White [R]			8.013	2	.018 [Sig]		
Race – Black	.089	.121	.540	1	.462	1.093	.863 – 1.384
Race – Hispanic	.234	.128	3.364	1	.067	1.264	.984 – 1.623

Note. N = 3,091. [R] indicates the reference group, LoS = Level of Supervision, and DRAOR = Dynamic Risk Assessment for Offender Re-Entry (Serin, 2007) where *B* is the regression coefficient, S.E. is standard error, and *df* denotes the degrees of freedom for the analysis, and CI = confidence intervals. [Sig] indicates that the reference category of the covariate reliably predicts survival and * denotes significance for comparisons between the reference group and other levels of a covariate; alpha values of 0.006 was used to adjust for inflated familywise error rate.

Discussion

The purpose of Study 2 was to explore whether rates of recidivism and survival time differed for White, Black, and Hispanic JI women in Iowa. Rates of technical violations, new offences, and any return (base rates) were calculated for each race group and compared using Chi-square tests. Cox regression survival analysis was used to examine differences in failure rates by race, controlling for level of static and dynamic risk. Significant differences in base rates and survival time were identified across race groups, though, in several cases, the observed differences were inconsistent with hypotheses.

Summary of Findings

Differences in base rates. Base rates of technical violations, new offences, and any return were calculated by race alone and by race at each level of the static (LoS) and dynamic (DRAOR scores) risk variables. Black women were significantly more likely than their White and Hispanic counterparts to incur a technical violation while on supervision but were not more likely to commit a new offence. Thus, the hypothesis that base rates would be higher for Black women was only partially supported. Hispanic women were least likely to experience any of the three types of recidivism examined while on parole, though this difference did not always reach statistical significance.

When base rates were further disaggregated by LoS and DRAOR scores, several concerning patterns emerged. First, for technical violations, the pattern of base rates observed for Black women did not increase monotonically as LoS and DRAOR scores increased (see Tables 15 and 16). Results demonstrated that Black women supervised at a

low normal LoS incurred technical violations at the same rate as those supervised at the intensive level, and that women in at the high normal LoS had even higher rates. However, for White and Hispanic women, increases in base rates appeared to mirror increases in LoS. Similarly, when base rates were disaggregated by race and DRAOR scores, the proportion of Black women with technical violations did not correspond to level of dynamic risk as would be expected. Black women in the low moderate, moderate, and moderate high dynamic risk groups were found to incur violations at virtually the same rate (range = 71.7% - 74.5%). For reference, the base rate for White and Hispanic women in the low moderate DRAOR category was roughly 50%. Also noteworthy, White women in the highest risk DRAOR category had a base rate of 56.0% for technical violations. This is surprising given that (a) this rate approximates that of the lowest risk group for White women and (b) the base rate for the next highest risk group (moderate high) was 72.6%.

With respect to new offences, observed patterns of base rates were consistent with expectations with one exception. White women with the lowest DRAOR scores (the low moderate group) were more likely (though only slightly) to commit a new offence than White women assessed as moderate risk (9.7% versus 9.3%). The pattern of base rates observed for the any return outcome paralleled those seen for technical violations. Clearly, there is a need to more carefully examine the routine supervision practices employed with women in general, and with Black women and with high and low risk White women.

Differences in survival time. Hypotheses with respect to survival time were also partially supported. Results provided some evidence that survival time was associated

with race, but only for the technical violations and any return outcome. Race had no meaningful impact on likelihood of committing a new offence. Equally, findings regarding the relationship between LoS and DRAOR scores (i.e., static and dynamic risk) and survival were mixed. While the LoS covariate was consistently related to survival time for all three outcomes examined in this study, DRAOR scores did not have a reliable effect on time to new offence. Notable results for each outcome are summarized below.

Technical violations. Being Black significantly increased odds of failure compared to being White (*OR* = 1.167, *p* = .005). LoS was also significantly associated with time to failure, with higher levels predicting shorter time to failure; women supervised at high normal and Intensive levels failing at 3.4 and 3.9 times the rate of women supervised at the administrative level. Considered collectively, DRAOR scores were associated with survival time, but the odds of failure did not increase proportionately with DRAOR scores. Contrary to expectations, the highest risk group had the lowest odds of failure. Though the difference in survival time between the high risk women and women in the low moderate group (the reference group) was not statistically significant (*OR* = .826, *p* = .400) this result was especially surprising considering that the women with moderate high scores were more likely to fail than those in the low moderate group (*OR* = 1.251, *p* = .001).

New offence. Race did not have a reliable effect on survival time for the new offence outcome variable. Survival time was shortest for White women and longest for Hispanic women, but differences in time to failure were not statistically significant. LoS was the only covariate that did predict survival time for this outcome, with significant differences between the failure rates for women who received intensive supervision and

women who received minimum or administrative supervision. Importantly, women in the administrative group had a 100% survival rate – that is, none committed a new offence while in the community.[18] New offence rates for the minimum LoS group were also incredibly low (i.e., $n < 10$). Thus, significant differences in survival time were unsurprising. When compared to women supervised at the low, and high normal levels, differences in survival time were not significant. DRAOR scores did not reliably predict time to new offence nor were differences between the survival curves for each DRAOR risk group significant. Nevertheless, results from this analysis warrant discussion. Survival time appeared to be positively correlated with risk, rather than negatively. In other words, the highest risk women also evidenced the longest survival time and the purportedly lowest risk women failed most quickly. Again, though not statistically significant, these results raise questions as to the supervision practices utilized with JIPs at ends of the dynamic risk continuum.

Any return. Any return is commonly used as an index of recidivism in correctional research and was therefore examined in this study despite clear limitations to utility. For instance, in the current study, of the 2,126 events comprising the any return outcome, 2,074 (97.6%) are technical violations. As such, new offences contribute very little information to this outcome and do little more than distort the findings based solely on technical violations. Accordingly, findings in this section should be understood as such. Race, LoS, and DRAOR scores all had an effect on survival time at the level of the

[18] The intensive supervision group was used as the reference category for the new offence outcome because the base rate for new offences in the administrative group (used as the reference group for the technical violations and any return outcome) was 0; using the administrative group as the reference category would have significantly biased comparisons of survival time and odds ratios.

covariate (i.e., including each of the risk and race covariates in the model improved prediction), but differences in survival time according to group within the covariate (i.e., being White, Black, or Hispanic within the race covariate) were not especially meaningful. It is recommended that readers refer to the technical violation and new offence outcomes separately for a more accurate and nuanced representation of the impact of the risk and race covariates on survival time.

Contextualizing the Findings

The recidivism rates observed in this study are high compared to the base rates typically reported by the Iowa Department of Corrections (IDOC) and previous research using Iowa community samples. In this sample, 67.1% of women incurred a technical violation and 10.6% committed a new offence for an overall return rate of 68.8%; recidivism rates cited in other sources range from approximately 30-50%. For example, in 2013, Prell conducted a validation study of the Iowa Risk Assessment Revised using a sample of 1,961 male and 2,006 female JIPs being supervised on probation or parole in Iowa. She reported recidivism rates (any new crime and/or revocation) of 41.6% and 29.9% for men and women, respectively. Similarly, in their Prison Recidivism Report for fiscal year 2016, the IDOC reported a return rate of 30.0% for JI women. Moreover, this report also indicated that there was no statistically significant difference in recidivism between Whites and Blacks. Following a concerted effort to minimize disparities in community failure through increased reentry supports for Black JIPs, the recidivism rate for both Black and White JIPs (men and women), was 35%. In this study, the return rate for Black women exceeded 75%.

The difference in observed recidivism rates can be potentially explained by differences in risk level. The women in this study were high risk. Prell (2013) indicated that 10% of the women in her sample were received Intensive supervision, and 12% were supervised at the high normal level. The remaining 78% were classified as low normal, minimum, or administrative. By comparison, 39.4% of the women in this study required Intensive supervision and 28.9% were supervised as high normal. In other words, 68.3% of this sample were considered high risk.[19] Notably, this disparity was even more pronounced when looking at Black women, 72.5% of whom were supervised at either the high normal or intensive level.

Chadwick (2014) conducted a validation study of the DRAOR using a sample of 391 (20.5% female) Iowa probationers and parolees. Like the current study, his sample also contained a large proportion of high risk JIPs, with the average JIP being assigned to Intensive supervision. Chadwick (2014) reported a base rate of 47.8% for any recidivism. Though considerably higher than the recidivism rates described above for the average Iowa JIP (i.e., non-high risk JIPs), this rate is still much lower than the rates observed for JI women in the current study.

Possible Explanations of Findings

Moving beyond concerns regarding the elevated base rates seen for all women in this study, two main lines of questioning remain. The first pertains to the pattern of base rates and survival time seen for Black women in relation to the technical violations and any returns outcome and the second relates to the unexpected findings with respect to the

[19] The DRAOR is not applied to low risk offenders. Thus, it is expected that the sample would be comprised of offenders assessed as higher risk rather than low risk.

survival time of high and low risk White women. Possible contributing factors to these outcomes are discussed below.

For Black women in particular, rates of technical violations did not appear to vary as a function of static and dynamic risk, which is concerning. Typically, lower risk women would be expected to incur fewer violations, which was not the case here. Possible explanations for this include: (a) lower risk Black women are receiving many conditions of supervision (some of which may not be necessary) and are committing violations as a result; and (b) the day-to-day realties of Black women may present considerable obstacles their reintegration efforts, beyond those experienced by non-Black women.

The nature, scope, and utility of supervision conditions has been a topic of research for more than fifty years, beginning with the seminal work of Arluke (1956, 1969). Conditions are intended to provide guidance to the JIP and ultimately facilitate his or her successful reentry into society, thereby striking a balance between the goals of offender rehabilitation and public safety (Skeem & Manchak, 2008). However, many (e.g., Arluke 1956, 1969; Jacobson 2005; Petersillia 2001; Solomon et al., 2008; Travis & Latessa, 1984; Travis & Stacey, 2010) have noted the liberal assignment of conditions with concern. Travis and Stacey (2010) describe these concerns succinctly:

> More numerous and comprehensive conditions increase the probability that a parolee will violate at least some constraint…if we seek to reduce the rate of "technical" failures on parole, we might do well to minimize the number and extent of parole conditions (p. 607)

The prevailing opinion appears to be that less is more when it comes to conditions. Hanlon, Nurco, Bateman, and O'Grady (1999) offer a reminder that this is especially true when it comes to numerous standard conditions as they may not be appropriate (or effective) when assigned to JIPs with differing needs. Thus, conditions of supervision may be functioning as a double-edged sword for Black women, especially for those who are lower risk or who have numerous, generic conditions.

Equally, understanding of an JIP's re-entry context is central to understanding recidivism. There is now considerable research suggesting that women face unique challenges during the reintegration process (see, for example, Belknap, 2007; Bloom et al., 2003; Chesney-Lind, 1997; Daly, 1998; Daly & Chesney-Lind, 1988; Miller & Mullins, 2006; Owen, 1998). In addition to pre-imprisonment and demographic characteristics such as younger age, being incarcerated for a property offence, previous substance use, and lengthy criminal histories that have all been shown to meaningfully influence recidivism (e.g., Kruttschnitt & Gartner, 2003, 2005; Stuart & Brice-Baker, 2004), aspects of re-entry context (e.g., a return to substance use, impoverished neighbourhood) are increasingly being implicated in community supervision failures (Deschenes, Owen, & Crow, 2007). Many women, and particularly women of colour, return to disadvantaged neighbourhoods (Dodge & Progrebin, 2001; Owen & Bloom, 1995; Richie, 2001) and are, correspondingly, more likely to experience difficulties accessing programs and services. Neighbourhood context can also differentially affect women by reducing the likelihood that they will be able to find meaningful employment. Other aspects of prosocial life, including prosocial friendships and marriage to partners with little or no prior criminal involvement are also adversely affected (Edin & Kefalas,

2005). Finally, finding safe, stable housing, recognized as an essential element of re-entry success (e.g., Petersilia, 2003; Sullivan, Mino, Nelson, & Pope, 2002; Visher & Courtnery, 2006) is considerably more challenging for women returning to impoverished neighbourhoods. Coupled with the emphasis currently placed on "treatment" focused conditions (e.g., pursuing education, maintaining employment, and completing drug, alcohol, and/or psychological programs; Travis & Stacey, 2010), it seems reasonable to expect that rates of technical violation would be higher among Black women.

With respect to explaining why lower risk women (both Black and White) may incur more violations and/or experience shorter time to failure when compared to their purportedly higher-risk counterparts, several possibilities also exist. For instance, parole officers may prioritize higher risk women in such a way that lower risk women are not receiving the support they need. There may also be other underlying issues related to the appropriateness of the IVVI and DRAOR for use with JI women which could impact how scores are being used to inform supervision practices.

With respect to the former, parole officers may be inclined to prioritize JIPs assessed as higher risk. Many parole officers are often pressed for time (e.g., Adelman, 2020; Matz et al., 2018) and are faced with decisions about how to best divide their time and direct their efforts among JIPs on their caseload. When making such decisions it is reasonable to expect that, in accordance with the Risk principle (see Andrews and Bonta, 2010 for more information about the Risk, Need, Responsivity model) that parole officers would opt to prioritize their highest risk clients. Concentrating on high risk clients might increase the likelihood of recidivism for lower risk clients if their risk and need factors are not adequately mitigated; low risk does not mean zero risk (i.e., 100 percent success).

Equally, it is possible that the IVVI and DRAOR are simply not well-suited for use with JI women , and Black women in particular. A central goal of this dissertation was to explore how gender and race might interact to disadvantage JI women belonging to historically marginalized groups, and that may be a contributing factor here. Though both the IVVI and DRAOR have been tested and validated in samples including JI men and women (e.g., Pardoel, 2020; Prell, 2013; Serin, Chadwick & Pardoel, 2018), there is little research examining how these instruments function for women belonging to historically marginalized groups, and no published studies have explored this when further disaggregating by risk level. Moreover, in Study 1, the DRAOR did not demonstrate measurement invariance across race. That is, a score of 23 may not have the same implications for a Black woman and a White or Hispanic woman. These findings suggest that when the DRAOR is used to assess JI women, careful consideration of how scores are interpreted and used to inform practice is paramount. In the context of the current study, what this means is that scores, and changes in scores on individual items may be more important than overall scores and category labels in terms of effective case management. As discussed above, re-entry context is also critical for women, especially marginalized women. Thus, women with lower DRAOR scores and/or a lower LoS might not actually be at lower risk for recidivism if certain unaccounted for risk factors and particular circumstances are present in her life. Consider for example, two hypothetical women. The first has a DRAOR Total score of 12, based on the presence of a moderate amount of risk factors, counteracted by many protective factors. The second has a DRAOR Total score of 6, based on few risk factors, but also few protective factors. The first woman returns to live with her prosocial parents following release and has a family

member who is prepared to give her a job. The second is a single mother returning to a poor neighborhood with limited employment prospects and access to services. Clearly, re-entry context cannot be ignored.

Implications for Practice

Results of Study 2 have considerable implications for routine supervision practice of women clients in Iowa. First and foremost, parole officers should be aware that reliance on total scores or classifications determined using traditional risk assessment measures to inform supervision strategies is not advisable for JI women, and JI Black women in particular. To be clear, this is not to say that static and dynamic assessment tools such as the IVVI and the DRAOR are without value when it comes to informing aspects of supervision. On the contrary, these tools provide essential information that can and should be used to tailor supervision to each individual's risks and needs. However, total scores and supervision categories were not reliably associated with recidivism in this study and parole officers should therefore be careful to not use membership in a certain risk category as sole justification for a particular approach. Instead, a recognition that risk of recidivism is much more nuanced for JI women is required. To mitigate risk, supervision strategies should aim to address each woman's unique risks and needs and carefully consider how they might interact with her immediate re-entry context.

Limitations and Future Directions

Results in Study 2 should be considered in light of several limitations. First, how the outcome variables (i.e., technical violations, new offence, and any return) were defined was quite broad and may have contributed to imprecision in base rates and analyses related to survival time. Technical violations were defined as any behaviour that

violates one or more conditions of parole. Thus, violations ranged from minor (e.g., failing to abide by a curfew or missing a meeting with their parole officer) to serious (e.g., violating a no contact order or possessing or purchasing dangerous weapons). Similarly, new charges varied considerably, ranging from petty crimes (e.g., petty theft or minor acts of vandalism) to serious violent offences. The any return outcome variable is also inherently flawed; technical violations account for the vast majority of events captured by this variable, and the inclusion of a much smaller number of new offences likely does little more than muddy the waters. Future research should seek to incorporate more precisely defined outcome variables to better capture event severity; distinguishing between more serious forms of recidivism and less correctionally relevant supervision outcomes would increase accuracy and improve the utility of findings. The current study is also limited by the fact that it relied on archival data. As is common when working with archival datasets, the dataset made available by the Iowa Department of Corrections for this dissertation was lacking a number of potentially informative explanatory variables such as staff characteristics or information about what is discussed during supervision meetings .

Despite apparent sample size, sample composition was also a limitation in this study. Black women (n = 560) represented only 18.1% of the overall sample, and Hispanic women (n = 113) accounted for less than 4.0%. Larger sample size for these groups would increase options with respect to possible analyses (e.g., disaggregating by race and risk simultaneously to examine survival curves) and may have increased the clarity and precision of results. Furthermore, power in Cox regression survival analysis is affected by not only overall sample size, but also by equality of sample size (Tabachnick

& Fidell, 2013). The strength or magnitude of the effect of the predictor variables included in the analysis also has an impact on power. In the present study, the covariates (LoS and DRAOR scores) were found to be weakly associated with survival time. Thus, in the absence of strong effects of the covariates, maximizing power via other means would be desirable.

In addition to addressing these limitations, future research should seek to clarify why the base rates observed in this sample were so high. As discussed above, how the outcome variables were defined in this study are likely a contributing factor, however, it is unlikely that this is the only factor. Previous research conducted in Iowa has reported base rates between 30% and 50%, even when samples included high-risk men. As such, determining whether this sample is representative of the current population of JI women in Iowa or an anomaly would appear to be an important task. Factors that have been demonstrably associated with increased likelihood of recidivism in women (e.g., age, prior substance use, proportion of women with a property or drug offence; Deschenes, et al., 2007; Huebner, DeJong, & Cobbina, 2010; Kruttschnitt & Gartner, 2003, Staurt & Brice-Baker, 2004) represent potential starting points. Relatedly, it is recommended that future research examine the type and number of conditions assigned to JI women, and Black women in particular, to examine how conditions might be related to technical violations. Finally, a review of how supervision strategies are established and adjusted for JI women based on assessment scores is recommended. Increased understanding of what information parole officers prioritize (i.e., overall scores or categories, domain scores, individual item scores) with their female clients and how it translate to aspects of their supervision approach could contribute to the broader field of offender reintegration.

Conclusion

The current study confirms that correctional agencies have an obligation to consider the effect of gender and race on the reintegration process and post-release success. The recidivism rates observed in this sample were especially high, especially for Black women. This points to a need for additional research to examine the factors underlying technical violations and new offences among JI women, and increased attention to women's re-entry context is advised. Findings regarding the lack of variability in recidivism rates for Black women supervised in Iowa across levels of static and dynamic risk also warrant further investigation. With respect to the supervision of JI women more broadly, further research exploring items that might function as flags of imminent recidivism is suggested. Results of the current study warn against reliance on risk categories and assessment total scores and instead suggest that mitigating risk for JI women might be best accomplished by identifying and addressing case-specific risk factors. Research aimed at identifying which factors might be most salient for women, taking into account important factors such as race and re-entry context, is recommended.

Chapter 7: Study 3 – Exploring Prediction

Purpose

Study 3 aimed to explore the predictive abilities of the DRAOR while attending to intersectionality. Justice involved women are widely recognized as one of the fastest growing correctional populations (e.g., Blanchette & Brown, 2006; Brown, et al., 2017; Greiner, et al., 2014; Guerino et al., 2011; Public Safety Canada, 2016), and therefore

merit increased attention in correctional research. Equally, the extent of overrepresentation of historically marginalized groups in prison and the sheer size of North American correctional populations necessitates improvements to risk assessment practices with JI women and historically marginalized groups. Seeking to clarify how different risk and protective factors (and by extension, overall risk assessment scores) might differentially impact JI women with different racial backgrounds is a logical starting point as findings could be used to improve the efficacy of case management practices for this population. For instance, a better understanding of which item(s) or subscale score(s) are especially salient for JI Black and Hispanic women could aid parole officers in the development and timely revision of intervention plans, thus improving outcomes for these women. Correspondingly, more definitive findings with respect to the ability of the items and subscales comprising the DRAOR to predict various types of recidivism for JI women with different racial backgrounds has clear implications for case management and could potentially aid in the amelioration of the pervasive system-wide issues like racial disproportionality. The analyses undertaken in Study 3 sought to explore these predictive nuances and, to better understand the impact of gender, utilized matched samples of JI men and women.

Importantly, to comprehensively assess the predictive accuracy of tools such as the DRAOR, both relative and absolute risk must be considered (Hanson, 2017, Helmus & Babchishin, 2017; Singh, 2013) These two aspects of predictive accuracy are complimentary, but distinct. Relative risk – also known as discrimination – allows for consideration of one JIP in relation to another (Hanson, 2017). Thus, it is the extent to which a risk assessment tool accurately distinguishes recidivists from non-recidivists

(Hanson, 2017; Helmus & Babchishin, 2017; Singh, 2013). Conversely, absolute risk – also referred to as calibration – provides the expected probability of recidivism associated with a given score, and absolute predictive accuracy evaluates the degree of concordance between the recidivism rates predicted and the actual rates observed in a new sample (Hanson, 2017; Helmus & Babchishin, 2017). For example, a scale is well calibrated if it predicts that 35% of JIPs with a given score will reoffend and subsequent validation studies find recidivism rates close to 35% for JIPs with that score. The first part of Study 3 is concerned with assessing the discrimination and calibration of DRAOR Total and subscale scores for White, Black, and Hispanic JI men and women.

The second part of Study 3 focused on exploring predictive nuances at the item level. To date, research on the DRAOR has not explored the predictive abilities of individual items while disaggregating by gender and race. However, existing correctional research increasingly supports the idea that many dynamic risk and protective factors may be more or less salient for JI women (e.g., Jones et al., 2015; Lodewijks et al., 2010) and that some protective factors may have a promotive, or even protective, influence.[20] Thus, in addition to exploring the predictive ability of all DRAOR items in relation to technical violations, new offences and any return, a key focus in the second half of Study 3 was to examine the Protective subscale and to determine if any of the items included in this domain demonstrated promotive or protective effects for JI women.

[20]As a reminder from Chapter 2, this dissertation adopts the conceptual framework put forth by Jones et al., (2015), in which promotive factors are defined as variables that are negatively correlated with criminal behaviour, regardless of an offender's risk level, and protective factors are understood to interact with risk level, such that they are more salient for higher risk offenders.

Two central hypotheses were tested in this study:

Hypothesis 1. It was expected that DRAOR Total and subscale scores would display adequate predictive validity for both male and female JIPs, regardless of race. Previous research on the DRAOR (see Chapter 3) provides evidence for moderate discrimination (*AUCs* of roughly .60) across various subpopulations, including men, women, youths, violent offenders, sex offenders, and Indigenous (Maori) offenders. Moreover, recent research by Serin, Chadwick, and Pardoel (2018) provided preliminary support for both the relative and absolute predictive accuracy of DRAOR Total scores in a sample of racially diverse JI men and women in Iowa. Overall, the results of this examination suggested that DRAOR Total scores modestly discriminated between individuals who experienced technical violations, serious violations, and new charges and those who did not. The relative predictive accuracy of DRAOR Total score was similar for technical violations and serious violations (*AUCs* = .64 and .63, respectively), and slightly smaller for new charges (*AUC* = .59). Predictive accuracy was also largely comparable across the race and gender subgroups, with no discernible pattern of differences emerging. Collectively, findings suggested that discrimination was adequate for men and women, and among White, Black, and Hispanic individuals though difference tests did reveal that predictive accuracy was elevated for White individuals compared to Black individuals when looking at technical violations and serious violations. Notably, Serin et al., (2018) were not able to disaggregate by gender and race simultaneously.

Serin and colleagues (2018) also found that the DRAOR demonstrated excellent calibration for White, Black, and Hispanic JI men and women. The absolute predictive

ability of DRAOR Total scores was largely consistent across JIP subgroups, and despite slight differences in the variability of calibration indices, findings suggested that DRAOR Total scores were able to predict recidivism accurately regardless of individuals' gender and race. In view of these findings, DRAOR scores were expected to demonstrate similar predictive accuracy (both relative and absolute) across the six gender by race groups examined in this study.

Hypothesis 2. Differences in item-level prediction were expected across gender, but not race. Research by Olver et al. (2014) on the Level of Service tools suggests that both substance use and personal/emotional factors are especially salient for women, and many others have also found associations between substance use (e.g., Cimino et al. 2015; Gobeil et al., 2016; Kopak et al., 2015) and adverse emotional states (e.g., Benda, 2005), and recidivism in women. Also, as noted by Brown (2017), women consistently score higher on certain risk/need domains such as dysfunctional relationships, criminal intimate partners, and housing concerns. As such, the DRAOR items related to these domains (e.g., substance use, negative mood, relationships, employment, etc.) were expected to be especially predictive for women in the sample.

To date, DRAOR research has not examined whether the predictive utility of individual items differs as a function of race. Findings from recent research based on other assessment tools (e.g., the PCRA; Skeem & Lowenkamp, 2016) suggest that variation in static factors (i.e., criminal history) across racial groups accounted for the vast majority of observed variation in risk scores, whereas variation in dynamic factors made a negligible contribution. In view of this, DRAOR items were expected to demonstrate equivalent predictive abilities across racial groups.

With respect to promotive and protective factors, items in the Protective subscale were expected to function as promotive factors, if not protective factors for JI women. No a priori hypotheses were made with respect to differences in promotive function across race. Also, in view of the stringent definition of protective factors espoused in this dissertation, identification of true protective factors, while possible, was not expected. There were no a priori hypotheses regarding particular candidate items.

Method

Participants

Study 3 utilized matched samples of men (n = 2,763) and women (n = 2,763) JIPs serving community supervision orders in Iowa. To control for the effects of age, risk, and race on recidivism, the case control procedure in SPSS was used to create matched samples.[21] Sample equivalence was confirmed using paired samples t tests for the continuous age and DRAOR score variables, and a chi-square test for independence for the categorical race variable. Mean age and DRAOR score were not significantly different for men and women, with $t(2762) = 1.88$, $p = .091$ and $t(2762) = -1.51$, $p = .113$, respectively. Likewise, the proportion of White, Black, and Hispanic JIPs did not differ between the male and female sample ($\chi^2(2) = 2.91$, $p = 2.33$).

Table 20 summarizes key sample characteristics disaggregated by gender and race. As age and risk were controlled for when creating the samples, mean age and DRAOR Total scores show little variability across the groups. Base rates, however, do

[21] Sample size in this study was limited to the number of cases that could be matched on age, risk, and race. Increasing the number of controls (i.e., marital status, education, DRAOR subscale score) significantly decreased available sample size which was not desirable.

vary across gender and race groups. Overall, women were more likely than men to incur technical violations (67.3% versus 62.0%) while men were more likely to commit a new offence (13.6% versus 10.7%). Across both genders, Black JIPs had the highest rate of any recidivism.

Table 20

Age, Dynamic Risk Assessment for Offender Re-Entry scores, and Recidivism Rates for Matched Samples of JI Men and Women

	White *n* (%) | *M* (*SD*)	Black *n* (%) | *M* (*SD*)	Hispanic *n* (%) | *M* (*SD*)
Men (*n* = 2,763)	2,194 (79.0)	491 (17.8)	78 (2.8)
Age	35.7 (9.6)	34.1 (9.8)	32.1 (7.6)
DRAOR score	6.9 (6.9)	7.5 (6.2)	5.9 (6.0)
Rate of Technical Violations	1,358 (61.9)	302 (61.5)	52 (66.7)
Rate of New Offence	279 (12.7)	86 (17.5)	11 (14.1)
Rate of Any return	1408 (64.2)	322 (65.6)	53 (67.9)
Women (*n* = 2,763)	2,167 (78.4%)	496 (18.0%)	100 (3.6%)
Age	35.0 (9.8)	32.8 (10.4)	32.1 (9.5)
DRAOR score	7.2 (7.3)	7.3 (6.8)	7.2 (7.2)
Rate of Technical Violations	1,427 (65.9)	368 (74.2)	65 (65.0)
Rate of New Offence	237 (10.9)	52 (10.5)	6 (6.0)
Rate of Any return	1,464 (67.6)	376 (75.8)	67 (67.0)

Procedure, Measures and Outcome Data

The data collection procedure, relevant measures, and outcome variables examined in this study are described earlier in this dissertation (see Chapter 4). Notable differences for this study include: (a) the inclusion of a matched sample of JI men (described above); and (b) utilization of a set of comparison (or reference) recidivism

rates drawn from data collected during an earlier pilot study of the DRAOR in 2014 (Chadwick, 2014) for calibration analyses. Relevant aspects of this dataset are presented in the section describing calibration analyses.

Analytic Approach

Analyses for Study 3 were conducted using SPSS (version 25), Microsoft Excel, and the pROC statistical package for R (Robin et al., 2011).

Mean differences by gender and race. Prior to exploring the predictive accuracy of DRAOR subscales and items, means were calculated for each gender and race group to provide a frame of reference and facilitate understanding of subsequent item-level analyses. For instance, if Black women consistently received high scores on a given item and this item then demonstrated high predictive accuracy, this information could be used to inform supervision practices.

Mean differences in DRAOR variables across gender were assessed using a series of independent samples *t* tests. Cohen's *d* effect sizes were also calculated for each comparison. As multiple comparisons were made, it was necessary to control for inflated Type I error rates. The Benjamini-Hochberg False Discovery Rate (FDR) method was selected over the Bonferroni method for its ability to control for inflated Type I error while preserving power (Benjamini & Hochberg, 1995). This method is less likely to produce false negatives (i.e., to discard significant observations) as it takes the relative ranking of *p* values into account rather than uniformly increasing the stringency of the criteria for significance across all comparisons. Using the Benjamini-Hochberg FDR approach, all achieved *p* values are placed in descending order before being divided by their position in the overall order and multiplied by the desired alpha value. Proceeding in

order, the significance of individual *p* values is assessed. Once the first significant value is identified, all subsequent (i.e., smaller) *p* values will also be significant. Note that this approach was also used to control for inflated Type I error rates in later analyses in Study 3 involving multiple comparisons. Effect size was calculated and evaluated according to Cohen's (1992):

$$d = \frac{M1-M2}{SD\ Pooled} \text{ where } Pooled\ SD = \sqrt{\frac{(SD1^2+SD2^2}{2}}$$

Effect sizes of 0.2 were considered small, 0.5 considered moderate, and 0.8 considered large (Cohen, 1992).

Assessing predictive ability. The predictive ability of DRAOR Total and subscale scores was assessed by means of discrimination and calibration analyses. Analyses were conducted separately for each race group within gender using the case-matched samples of men and women. To date, evaluations of the predictive of accuracy of risk assessment instruments have relied primarily on discrimination. While there are a number of appropriate statistics for tests of discrimination (i.e., area under the curve [*AUC*s], Cox regression, Harrell's *C,* Cohen's *d,* and logistic regression; Helmus & Babchishin, 2017), there is no consensus regarding appropriate tests statistics for calibration, and little available guidance on evaluating absolute predictive accuracy in similar contexts (Hanson, 2017; Helmus & Babchishin, 2017). At present, options for evaluating absolute predictive accuracy are limited to Pearson's chi-squared goodness-of-fit statistic (see Doren, 2004; Harris, Rice, & Cormier, 2002; Harris et al., 2003b), and the E/O index (see Helmus & Thornton, 2016; Olver & Sewall, 2018).

Assessing discrimination. Receiver operating characteristic (ROC) analyses are one of the most commonly used assessment techniques to evaluate the predictive accuracy of assessment tools with respect to recidivism and were employed in this Study. Receiver operating characteristic analyses assess both the specificity and sensitivity of a measure (Rice & Harris, 2005), and generate the Area Under the Curve (*AUC*) statistics. Specificity pertains to a measure's ability to correctly identify a non-recidivist, whereas sensitivity refers to a measure's ability to correctly predict recidivism, termed *hit rate* (Rice & Harris, 2005). The resultant ROC curve plots the hit rate (sensitivity) against false alarms (obtained by calculating 1-specificity; Craig & Beech, 2009). Area under the curve statistics describe the probability that a score on a measure drawn at random from one sample is higher than a score drawn at random from another sample (Harris & Rice, 2005). In the context of the present study, *AUC* statistics represent the probability that the DRAOR score of a randomly selected recidivist would be higher than the DRAOR score of a randomly selected non-recidivist. Possible *AUC* values range from 0 to 1, where *AUC* values of .50 indicate that prediction is equivalent to chance. For the sake of simplicity, *AUC*s are often positively coded, so that scores above .50 are used to represent strong predictors, rather than scores closer to 0. Usefully, *AUC*s correspond to effect size conventions, which facilitates interpretation. As per Rice and Harris (2005), *AUCs* ranging from .55 to .63 are analogous to small effect sizes, *AUCs* ranging from .64 to .70 are analogous to medium effect sizes, and *AUCs* of .71 or higher are analogous to a large effect. The binormal *AUC* is given by the following formula (see Hanley, 1988):

$$\mathrm{AUC} = \Phi\left(\frac{\mu_X - \mu_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right)$$

For this study, a binormal smoothing method was selected over a trapezoidal approach (i.e., nonparametric *AUC*) because it is more powerful and tends to produce a more accurate estimation of the *AUC* (Hanley & McNeil, 1982). Moreover, this approach is robust to deviations of the underlying assumptions of parametric tests; the parametric approach assumes that both the noise (negative) population and the signal (positive) population are normally distributed. As discussed by Hanley (1988), smoothed ROC curves are robust to even strong deviations from this assumption.

In the current study, descriptive 95% confidence intervals were also calculated for each *AUC* using a bootstrapping method (n = 25,000)[22] to increase the stability of results (Robin et al., 2020). Briefly, bootstrap is a computational resampling technique that mimics the process of randomly sampling from an assumed infinite population and allows you to calculate standard errors, construct confidence intervals, and perform hypothesis testing (Robin et al., 2020). Bootstrapping is widely recognized as an appropriate approach for controlling and checking the stability of results (Robin et al., 2020).

These analyses were supplemented by unpaired difference tests for uncorrelated (independent) *AUC*s to determine whether relative predictive accuracy differed significantly across gender and race groups for each of the three types of recidivism examined.

[22] There is no hard and fast rule regarding the minimum number of bootstraps that should be used for a given analysis. Generally, increasing the number of bootstraps improves the stability of the results. Thus, if there are no issues with computational power, there is no reason not to generate large numbers of samples as was done here.

The pROC package in R (Robin et al., 2011) was used to calculate the *AUC*s and subsequent difference tests, which were calculated using the following formula:

$$D = \frac{AUC1 - AUC2}{s}$$

In the above formula, *s* is the standard deviation of the bootstrap differences.

Assessing calibration. As reflected in numerous meta-analytic reviews (Campbell, French, & Gendreau, 2009; Hanson, Helmus, & Bourgon, 2007; Hanson, & Morton-Bourgon, 2009; Lofthouse, Golding, Totsika, Hastings, & Lindsay, 2017; Olver, Stockdale, & Wormith, 2009, 2014; Singh, Grann, & Fazel, 2011; Smith, Cullen, & Latessa, 2009; Viljoen, Mordell, & Beneteau, 2012) examinations of the predictive accuracy of risk assessment scales have typically focused on discrimination, not calibration. To date, little calibration research has been conducted and as noted earlier, there is little guidance available to researchers regarding appropriate statistics and methodology (Hanson, 2017; Helmus & Babchishin, 2017). Available options currently include the E/O index and Pearson's chi-squared goodness-of-fit statistic.

The chi-square goodness-of-fit statistic assess correspondence, or "fit" between expected and observed recidivism rates (Helmus & Babchishin, 2017). Degree of calibration is determined by the significance of the statistic; nonsignificant chi-square values indicate that the instrument is well calibrated as the observed and expected recidivism rates do not differ significantly. However, as noted by Helmus and Babchishin (2017), the goodness-of-fit statistic does not provide an effect size, thus limiting its utility in the broader research context. Like the goodness-of-fit statistic, the E/O index also measures the fit between expected and observed recidivism rates (Helmus & Babchishin,

2017). However, in addition to providing a measure of statistical significance, the E/O index also provides an effect size (Hanson, 2017; Helmus & Babchishin, 2017). Statistical significance is determined through the E/O index confidence intervals (CIs); CIs containing 1 indicate that expected and observed recidivism rates are not significantly different (i.e., good calibration; Hanson, 2017; Helmus & Babchishin, 2017). Effect size is determined through the magnitude of the statistic. For example, an E/O index of 2 suggests the scale predicts twice the number of observed recidivists (e.g., $10_{expected}/5_{observed} = 2$; Hanson, 2017). This ease of interpretation is advantageous in the communication of risk. This, along with its ability to simultaneous test for significance and provide an effect size, made the E/O index preferable to the chi-square test when evaluating the calibration of DRAOR Total and subscale scores in this study.

Calibration analyses were conducted using SPSS and Microsoft Excel. Importantly, computation of the E/O index requires the use of base rates obtained from an appropriate comparison sample, which ideally, would be the published and validated recidivism norms for the assessment tool. The recidivism rates from the reference sample are used to calculate the number of recidivists that would be expected in the study sample. As there are no empirically derived and validated norms for the DRAOR, data from an earlier pilot evaluation of the DRAOR in Iowa (Chadwick, 2014) was utilized for these purposes.

E/O indexes and 95% confidence intervals (CIs) were calculated for each gender by race group for each of the three recidivistic outcomes. The number of expected recidivists was obtained by multiplying the proportion of recidivists observed in the reference sample (i.e., the Iowa Pilot dataset) by the total number of JIPs with the

corresponding score in the current sample. For example, when considering the any charge outcome, 56% of men with a DRAOR Total score of 8 recidivated in the reference sample, which equates to a recidivism rate of .56. In the matched sample used in this dissertation (i.e., men, the any charge outcome), there were a total of 133 JIPs with a DRAOR Total score of 8. Multiplying 133 by .56 yields the number of expected recidivists, 74.48 in this case. The number of actual recidivists in the sample (the 'observed' observed recidivists) is then used to compute the E/O index. If, for example, there were 81 recidivists in current sample, we would divide 74.48 by 81, which would produce an E/O index of .920. Ninety five percent CIs were calculated in accordance with the recommendations of Hanson (2017) using the following formula:

$$95\%\ \text{CI for } \frac{E}{O} \text{ index} = \left(\frac{E}{O}\right) e^{\left(\pm 1.96\sqrt{\frac{1}{O}}\right)}$$

As described above, CIs containing 1 are indicative of good calibration (i.e., the expected and observed number of recidivists are not significantly different). Importantly, it should be noted that the 95% CIs calculated here do not take the number of expected recidivists into account, and that statistical power is instead determined as a function of the number of recidivists in the validation sample (Hanson, 2017).

Exploring predictive nuances. Logistic regression was used to explore the predictive ability of DRAOR items and subscale scores for each gender group disaggregated by race. Logistic regression models were also used to explore potential promotive and protective effects for items comprising the Protective subscale for JI women. Logistic regression allows for the prediction of a discrete outcome (i.e., recidivism: yes or no) using a set of predictor variables. Importantly, logistic regression

can accommodate any combination of dichotomous, discrete, and continuous predictor variables (Tabachnick & Fidell, 2013), making it ideal for the current analyses. Moreover, logistic regression is very flexible, or tolerant, with respect to assumptions. Specifically, no assumptions are made about the distribution of predictor variables; they do not need to be normally distributed, linearly related, or demonstrate equal variance within each group (Tabachnick & Fidell, 2013). That said, Tabachnick and Fidell (2013) recommend performing some basic checks of assumptions prior to analyses as multivariate normality and linearity among predictors may enhance power, which was done. Finally, as a series of logistic regression models were evaluated, the Benjamini-Hochberg FDR was used to correct *p* value thresholds; results are discussed in relation to whether or not they achieved significance after the use of the correction.

While the reporting of odds ratios is common practice when conducting logistic regression to facilitate the interpretation of coefficients, this approach can be problematic. In their paper, Uanhoro, Wang, and O'Connell (2019) describe the challenges inherent to interpreting logits or log odds (the coefficients produced by the regression model) as the effects of predictor variables and conclude that reliance on odds ratios is a flawed approach for quantifying and documenting the magnitude of the effect of predictor variables across models and studies. Thus, a reliable index of effect size should be presented alongside odds ratios (*OR*) for logistic regression analyses. Unfortunately, there appears to be little consensus regarding the best approach to calculating an R^2 for logistic regression (see Allison, 1999; Menard, 2000; or Mittlbock & Schemper, 1996 for more information about available approaches). For the current study, McFadden's ρ^2 was selected as an effect size based on the recommendations of Allison (1999) and

Tabachnick and Fidell (2013). McFadden's ρ^2 (Chanda & Maddalla, 1983) is a transformation of the likelihood ratio statistic intended to mimic an R^2, with a range of 0 to 1, and is given by the following formula:

$$\text{McFadden's } \rho^2 = 1 - \frac{\text{LL}(B)}{\text{LL}(0)}$$

where LL(B) is the log-likelihood of the full model and LL(0) is the log-likelihood of the constant-only model. The rationale for this model is that LL(0) plays a role analogous to the residual sum of squares in linear regression. As such, this formula corresponds to a proportional reduction in error variance and is sometimes referred to as a "pseudo" R^2. With respect to interpretation, McFadden's ρ^2 tends to be lower than R^2 for multiple regression (Tabachnick & Fiddell, 2013), with values in the .2 to .4 range considered highly satisfactory (Hensher & Johnson, 1981).

Results

Data Management and Assumptions Tests

Prior to beginning analyses, steps were taken to ensure that the data met all required assumptions. As the sample of men was unique to Study 3, it had not yet been examined. DRAOR Total, subscale, and item scores were screened to ensure that all values were within a valid range and that there was no missing data. Basic assumptions (linearity, homoscedasticity, normality, univariate and multivariate outliers, and multicollinearity) were also assessed. Bivariate scatterplots suggested that DRAOR variables were linearly related and no univariate outliers were identified in the male sample. Equally, the assumption of homoscedasticity was supported, as variability in scores for each variable was consistent across all values of the other variables. As in

Study 1, a statistical test of normality (i.e., Kolmogorov-Smirnov's test) was eschewed in favour of a visual inspection of histograms, probability plots, and skewness and kurtosis values. When sample size is large, Kolmogorov-Smirnov's test of normality is too conservative and often identifies departures from normality (Tabachnick & Fidell, 2013). Multivariate outliers were identified using Mahalanobis' distance. In total, there were 14 cases that exceed the critical value of 13.82 ($df = 2$, $\alpha = .001$) and these cases were examined individually. No meaningful pattern causing the outliers was identified and examination of influence values (i.e., Cook's distance) suggested that these outliers did not strongly impact the distribution and fidelity of data. Accordingly, these cases were retained. Multicollinearity, evaluated using bivariate correlations between DRAOR subscales, was not an issue as none of the inter-correlations suggested a perfect or near perfect relationship between the predictors.

As noted earlier, logistic regression is relatively free of restrictions. However, there are a number of practical issues that should be taken into account. First, while not strictly required, multivariate normality and linearity among the predictors may enhance power as a linear combination of predictors is used to form the exponent in the logistic regression equation (Tabachnick & Fidell, 2013). As such, histograms and Q-Q (quantile-quantile) plots were created to examine the distribution of the residuals (i.e., the errors between observed and predicted values). No significant deviations from normality were detected. Second, logistic regression assumes a linear relationship between continuous predictors and the logit transformation of the dependent variable. The Box-Tidwell approach (Hosmer & Lemeshow, 1989) provides a simple approach to testing this assumption. Essentially, this method involves adding interaction terms between each

continuous predictor (DRAOR Total and subscale scores in this case) into the regression model. If any of the interaction terms achieves significance then the assumption is violated. This assumption was tested separately for the technical violations, new offence, and any return outcomes; no violations were detected. Third, logistic regression, like all varieties of multiple regression, is sensitive to high correlations among other predictors (Tabachnick & Fidell, 2013). As described above (for men) and in an earlier section (the female half of the sample), multicollinearity was not a concern. Other potential limitations of logistic regression analysis (i.e., ratio of cases to variables, expected cell frequencies, and independence of errors) were considered and were not problematic for the current study.

Evaluating Mean Differences

Mean differences were examined as a preliminary to the main prediction analyses. While mean DRAOR Total scores were similar for JI men and women by design, it was unclear whether differences in mean DRAOR Total scores existed across the three race groups. Additionally, similar DRAOR Total scores cannot be equated with similar mean scores on individual subscales and items; thus, exploring potential differences in scoring patterns was of interest in order to better contextualize later findings.

Gender. Overall, women had slightly higher DRAOR Total scores (M = 7.2, SD = 7.2) than men (M = 7.0, SD = 6.8), though this difference was not significant. As shown in Table 21, women scored significantly higher than men on the Acute subscale (M = 7.1, SD = 2.9 versus M = 6.7, SD = 2.8) and Protective subscale (M = 6.0, SD = 2.8 versus M = 5.6, SD = 2.8), though the effect size in both cases was weak (Cohen's d < 0.2) suggesting that the mean differences were trivial (Cohen, 1992). Women also scored

higher than men on 11 of the 19 DRAOR items (denoted by negative *t* values), three of which (negative mood, interpersonal relationships, and social control) were associated with a small effect size. Anger and access to victims were the only items for which men scored significantly higher than women.

Table 21

Mean Differences Between JI Men and Women for DRAOR Subscale and Item Scores

Variable	*t*	*df*	*p* value	Cohen's *d*
Stable Subscale	-1.759	5521.249	.079	-0.05
Peer Associations	-4.640	5519.008	<.0001*	-0.13
Attitudes towards Authority	2.050	5520.895	.040	0.06
Impulse Control	-2.851	5523.488	.004	-0.08
Problem Solving	-4.667	5523.406	<.0001*	-0.12
Sense of Entitlement	2.500	5514.115	.012	0.07
Attachment with Others	.122	5521.654	.903	0.02
Acute Subscale	-5.831	5518.663	<.0001*	-0.16
Substance Abuse	2.724	5517.129	.006	0.08
Anger	8.019	5521.182	<.0001*	0.21[Sm]
Access to Victims	5.737	5505.420	<.0001*	-0.15
Negative Mood	-17.856	5523.135	<.0001*	-0.47[Sm]
Employment	-5.868	5520.716	<.0001*	-0.15
Interpersonal Relationships	-10.473	5521.966	<.0001*	-0.29[Sm]
Living Situation	-3.652	5523.916	<.0001*	-0.10
Protective Subscale	-4.754	5523.819	<.0001*	-0.13
Response to advice	-2.915	5523.446	.004	-0.08
Prosocial Identity	-2.218	5523.999	.027	-0.07
High Expectations	-2.099	5523.458	.036	-0.06
Costs/Benefits	-1.339	5522.245	.181	-0.03
Social Support	-4.425	5509.966	<.0001*	-0.12
Social Control	-8.222	5518.105	<.0001*	-0.22[Sm]

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry; the FDR method for multiple tests was used to correct the *p*-value threshold; * denotes significance; [Sm] indicates a small effect size. Negative *t* values denote a higher mean score for women. As Levene's test for homogeneity of variance was violated for several of the tests, the non-parametric equal variances not assumed *t*-test results were reported.

Race. Differences in DRAOR Total score were evident across race. Total scores were highest from Black JIPs (*M* = 7.4, *SD* = 6.5) and lowest for Hispanic JIPs (*M* = 6.6, *SD* = 6.7), with White JIPs falling in between (*M* = 7.0, *SD* = 7.1). White and Black, White and Hispanic, and Black and Hispanic JIPs were compared separately.

Table 22 provides the results of the subscale and item-level comparisons between White and Black JIPs. Black JIPs scored higher on the Stable subscale (*M* = 6.3, *SD* = 2.7) than their White counterparts (*M* = 5.9, *SD* = 2.7), although the *d* value of -0.14 suggests that this difference is negligible. Black JIPs also had higher mean scores on three DRAOR items (attitudes toward authority, sense of entitlement, and anger), which were associated with small effect sizes. White JIPs had higher mean scores on four DRAOR items, but only two (substance abuse and negative mood) were considered meaningful (i.e., $d > 0.2$).

Table 22

Mean Differences Between White and Black JIPs for DRAOR Subscale and Item Scores

Variable	*t*	*df*	*p* value	Cohen's *d*
Stable Subscale	-4.024	1489.149	<.0001*	-0.14
Peer Associations	-.949	1509.363	.343	-0.03
Attitudes Towards Authority	-6.471	1418.546	<.0001*	-0.23[Sm]
Impulse Control	1.454	1468.932	.146	0.05
Problem Solving	-.744	1534.950	.457	-0.03
Sense of Entitlement	-6.165	1415.132	<.0001*	-0.22 [Sm]
Attachment with Others	-2.112	1456.226	.035	-0.08
Acute Subscale	2.231	1562.618	.026	0.08
Substance Abuse	8.597	1408.340	<.0001*	0.32 [Sm]
Anger	-8.015	1436.399	<.0001*	-0.29 [Sm]
Access to Victims	-2.888	1461.270	.004	-0.11
Negative Mood	9.961	1495.394	<.0001*	0.35 [Sm]
Employment	-2.880	1481.893	.004	-0.11
Interpersonal Relationships	3.972	1489.958	<.0001*	0.14
Living Situation	-.847	1460.964	.397	-0.03
Protective Subscale	1.792	1528.192	.073	0.06
Response to advice	.193	1464.548	.847	0.01
Prosocial Identity	.122	1499.918	.903	0.01
High Expectations	1.371	1497.211	.171	0.05
Costs/Benefits	1.427	1500.858	.154	0.06
Social Support	3.737	1536.951	<.0001*	0.13
Social Control	1.026	1535.190	.305	0.03

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry; the FDR method for multiple tests was used to correct the *p*-value threshold; * denotes significance; [Sm] indicates a small effect size. Negative *t* values denote a higher mean score for Black offenders. As Levene's test for homogeneity of variance was violated for several of the tests, the non-parametric equal variances not assumed *t*-test results were reported.

Results of the comparisons between White and Hispanic JIPs are presented in Table 23. No mean differences were observed in subscale scores and there was only one

difference at the item level. Compared to their Hispanic counterparts, White JIPs were rated significantly higher on the negative mood item.

Table 23

Mean Differences Between White and Hispanic JIPs for DRAOR Subscale and Item Scores

Variable	*t*	*df*	*p* value	Cohen's *d*
Stable Subscale	.167	192.010	.868	0.01
Peer Associations	-.170	192.460	.865	-0.01
Attitudes towards Authority	-.782	191.089	.435	-0.06
Impulse Control	.747	191.155	.456	0.05
Problem Solving	.461	191.769	.645	0.04
Sense of Entitlement	-.412	191.596	.681	-0.04
Attachment with Others	.932	192.384	.352	0.08
Acute Subscale	1.866	190.885	.064	0.14
Substance Abuse	1.796	191.151	.074	0.14
Anger	-1.282	192.186	.201	-0.09
Access to Victims	-.693	190.753	.489	-0.05
Negative Mood	3.661	192.357	<.0001*	0.28[Sm]
Employment	2.365	192.782	.019	0.17
Interpersonal Relationships	.534	191.775	.594	0.04
Living Situation	.704	193.640	.482	0.05
Protective Subscale	.312	194.842	.755	0.03
Response to advice	.623	191.470	.534	0.05
Prosocial Identity	.109	194.841	.913	0.01
High Expectations	-1.192	192.331	.235	-0.09
Costs/Benefits	1.704	193.813	.090	0.13
Social Support	.209	193.180	.835	0.02
Social Control	-.072	194.037	.942	-0.02

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry; the FDR method for multiple tests was used to correct the *p*-value threshold; * denotes significance; [Sm] indicates a small effect size. Negative *t* values denote a higher mean score for Hispanic offenders. As Levene's test for homogeneity of variance was violated for several of the tests, the non-parametric equal variances not assumed *t*-test results were reported.

The final table in this section (Table 24) summarizes the results of the comparisons between Black and Hispanic JIPs. With the exception of the employment

item, no significant differences in DRAOR subscale and item scores were identified. Black JIPs received higher scores on this item and this mean difference was associated with a small effect size.

Table 24

Mean Differences Between Black and Hispanic JIPs for DRAOR Subscale and Item Scores

Variable	*t*	*df*	*p* value	Cohen's *d*
Stable Subscale	1.887	243.290	.060	0.16
Peer Associations	.248	242.717	.805	0.02
Attitudes towards Authority	2.111	249.216	.036	0.18
Impulse Control	.085	241.986	.932	0.01
Problem Solving	.746	236.778	.456	0.08
Sense of Entitlement	2.370	252.524	.019	0.18
Attachment with Others	1.808	249.769	.072	0.15
Acute Subscale	.889	230.430	.375	0.07
Substance Abuse	-2.136	251.308	.034	-0.18
Anger	2.388	251.966	.018	0.20
Access to Victims	.566	241.088	.572	0.05
Negative Mood	-.862	244.049	.390	-0.07
Employment	3.487	247.852	<.0001*	0.28[Sm]
Interpersonal Relationships	-1.192	242.112	.234	-0.10
Living Situation	1.042	255.163	.298	0.08
Protective Subscale	-.516	250.592	.606	-0.04
Response to advice	.502	244.095	.616	0.05
Prosocial Identity	.046	254.643	.964	0.01
High Expectations	-1.714	243.693	.088	-0.14
Costs/Benefits	.958	249.868	.339	0.06
Social Support	-1.415	242.455	.158	-0.11
Social Control	-.521	246.263	.603	-0.05

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry; the FDR method for multiple tests was used to correct the *p*-value threshold; * denotes significance; [Sm] indicates a small effect size. Negative *t* values denote a higher mean score for Hispanic offenders. As Levene's test for homogeneity of variance was violated for several of the tests, the non-parametric equal variances not assumed *t*-test results were reported.

Discrimination

The relative predictive accuracy (i.e., discrimination) of DRAOR Total and subscale scores was assessed via a series of smoothed Receiver Operating Characteristic (ROC) curves. The area under the ROC curve (*AUC*) statistic is a measure of diagnostic accuracy and represents the probability that a score on a measure drawn at random from one sample (i.e., scores of recidivists) is higher than a score drawn at random from another sample (i.e., scores of non-recidivists). ROC analyses were conducted using the pROC statistical package for R (Robin et al., 2011). Descriptive 95% confidence intervals were calculated using the bootstrapping method (n = 25,000). Analyses were conducted separately for each race group within gender and *AUC*s were generated for both DRAOR Total and subscale scores for each of the three recidivism outcomes. Differences across subsamples were tested using the unpaired difference test to determine whether the relative predictive accuracy was significantly different for one gender by race group compared to another. Overall, results did not support the hypothesis that DRAOR Total and subscale scores would display adequate predictive validity for JI men and women, regardless of race. Though DRAOR scores did demonstrate moderate discrimination for some of the study groups, results were not consistent across all combinations of gender and race or all recidivism outcomes.

Technical violations. The *AUC*s describing the relative predictive accuracy of DRAOR subscale and Total scores for each of the six race by gender subsamples are presented in Table 25 below. Notably, discrimination varied considerably across groups, although some patterns did emerge. DRAOR subscale and Total scores consistently predicted technical violations for White men and women, though prediction was

relatively weak. Results indicated that a randomly selected individual who committed a technical violation would have higher DRAOR scores than a randomly selected non-recidivist between 56% and 60% of the time. According to the guidelines set forth by Rice and Harris (2005) for interpreting effect sizes, this equates to a small effect. DRAOR subscale and Total scores poorly predicted technical violations for Black JIPs and Black men in particular. For Black men, *AUC*s ranged from .49 to .52; thus, prediction was analogous to chance. The *AUC*s were slightly higher for Black women (.53 – .55), though only scores on the Stable subscale could be classified as having a small effect. Notably, the 95% CIs associated with this *AUC*s for Black and Hispanic JIPs are much wider than those obtained for White JIPs. This important as wide CIs indicate that we have very little knowledge about the effect and that additional information is needed. A 95% CI is typically interpreted as indicating the range within which we can be 95% certain that the true effect lies. Thus, if we return to the *AUC* of .55 obtained for Stable scores predicting technical violations for Black women, the breadth of the 95% CI (.49 to .61) suggests that cautious interpretation is required. This proviso applies even more strongly to the findings regarding the predictive accuracy of DRAOR scores with Hispanic JIPs. DRAOR subscale and Total scores appear to discriminate well for Hispanic men, with Stable scores in particular providing excellent discrimination (*AUC* = .73). However, the breadth of the CI tempers enthusiasm somewhat. Stable scores also appear to offer moderate discrimination for Hispanic women (*AUC* = .63), although again, the associated CIs are worryingly wide.

Results of the *AUC* difference tests (Table 26) confirm that relative predictive accuracy varies by race. Acute and DRAOR Total scores were better at differentiating

recidivists from non-recidivists for White men compared to Black men and White women compared to Black women, with $D_{\text{Acute}} = 3.11$, $p = .002$ and $D_{\text{Total}} = 2.75$, $p = .006$ for White versus Black men, and $D_{\text{Acute}} = 2.05$, $p = .041$ and $D_{\text{Total}} = 2.02$, $p = .044$ for White versus Black women. Results also indicated that DRAOR Stable scores predicted technical violations more accurately for Hispanic men when compared to both White men ($D = -2.63$, $p < .01$) and Black men ($D = -3.26$, $p < .01$). Additionally, DRAOR Total scores provided better discrimination for Hispanic men relative to Black men. With the exception of the differences described above for White and Black women, no significant differences in discriminative abilities were identified for women of different races. Lastly, all difference tests between men and women of the same race were nonsignificant. Contrary to expectations, these findings suggest that discriminative ability did vary according to a JIP's race.

Table 25

Discriminative Validity of Dynamic Risk Assessment for Offender Re-Entry Scores by Gender and Race for Technical Violations

	Stable *AUC* [CI]	Acute AUC [CI]	Protective AUC [CI]	DRAOR Total AUC [CI]
Men				
White (*n* = 2,194)	**.57 [.55-.60]***	**.58 [.56-.60] ***	**.56 [.54-.59] ***	**.59 [.57-.62] ***
Black (*n* = 491)	.52 [.47-.58]	.49 [.43-.54]	.52 [.47-.57]	.50 [.46-.56]
Hispanic (*n* = 78)	**.73 [.61-.85]*****	**.56 [.42-.71]***	**.61 [.48-.74]***	**.67 [.53-.80]****
Women				
White (*n* = 2,167)	**.59 [.56-.62]***	**.60 [.57-.62]***	**.57 [.55-.60]***	**.60 [.58-.63]***
Black (*n* = 496)	**.55 [.49-.61]***	.54 [.48-.59]	.53 [.46-.59]	.54 [.48-.60]
Hispanic (*n* = 100)	**.63 [.51-.75]***	**.55 [.43-.67]***	.53 [.41-.65]	**.61 [.48-.73]***

Note. AUC = Area Under the Curve; CI = 95% Confidence Intervals; **bold** * indicates small effect size, ** denotes medium effect size, and *** denotes a large effect size.

Table 26

AUC Difference Tests for Dynamic Risk Assessment for Offender Re-Entry Scores by Gender and Race for Technical Violations

	Stable		Acute		Protective		DRAOR Total	
	D	*p*	*D*	*p*	*D*	*p*	*D*	*p*
Comparisons across Race								
Men								
White v. Black	1.77	0.077	**3.110**	**.002**	1.60	.111	**2.751**	**.006**
White v. Hispanic	**-2.626**	**0.009**	0.217	0.828	-0.669	.503	-1.193	.233
Black v. Hispanic	**-3.258**	**.001**	-1.052	.293	-1.291	.197	**-2.308**	**.021**
Women								
White v. Black	1.297	.195	**2.045**	**.041**	1.53	.126	**2.018**	**.044**
White v. Hispanic	-0.635	.525	0.785	.434	0.714	. 475	-0.034	.973
Black v. Hispanic	-1.221	.222	-0.225	.822	-0.075	.941	-1.018	.309
Comparisons across Gender								
White – Men v Women	-0.907	.364	-1.082	.279	-0.563	.574	-0.744	.457
Black – Men v. Women	-0.685	.493	-1.210	.226	-0.202	.840	-0.759	.449
Hispanic – Men v. Women	1.254	.210	0.155	.877	0.885	.376	0.717	.473

Note. D = coefficient for difference test, *p* represents probability of results using a two-tailed significance test and **bold** denotes a significant different test.

New offences. Table 27 summarizes the results of discrimination analyses for DRAOR subscale and Total scores for White, Black, and Hispanic men and women for the prediction of new offences. Considered collectively, DRAOR scores poorly predicted new offences, with *AUC*s ranging from .45 to .60. While several of the *AUC*s could theoretically be classified as small effects (i.e., between .55 and .63; Rice & Harris, 2005), the associated confidence intervals were often quite broad. Overall, discrimination appeared to be best for White men and Black women, as evidenced by *AUC*s of .55 or higher for Stable, Acute, and DRAOR Total scores. Interestingly, Stable, Acute, and Total scores appeared to better discriminate recidivists from non-recidivist for Black women when compared to Black men, though *AUC* difference tests indicated that this difference was not significant (see Table 28). Similarly, although discrimination appeared to be higher for White men compared to women for Stable, Acute, and DRAOR Total scores, these differences in relative predictive accuracy were likewise non-significant. Discrimination was especially poor for Hispanic JIPs; prediction was meaningfully better than chance for only Protective scores for men (*AUC* = .60), and Stable scores for women (*AUC* = .55).

Table 27

Discriminative Validity of Dynamic Risk Assessment for Offender Re-Entry Scores by Gender and Race for New Offences

	Stable *AUC* [CI]	Acute AUC [CI]	Protective AUC [CI]	DRAOR Total AUC [CI]
Men				
White (*n* = 2,194)	**.57 [.53-.61]***	**.55 [.51-.58]***	.54 [.50-.57]	**.56 [.53-.60]***
Black (*n* = 491)	**.55 [.47-.62]***	.54 [.47-.60]	.54 [.48-.61]	**.55 [.48-.62]***
Hispanic (*n* = 78)	.45 [.29-.61]	.51 [.32-.70]	**.60 [.44-.76]***	.50 [.33-.67]
Women				
White (*n* = 2,167)	.54 [.50-.58]	.51 [.48-.55]	.54 [.50-.57]	.54 [5.0-.57]
Black (*n* = 496)	**.59 [.51-.67]***	**.58 [.51-.65]***	.54 [.46-.62]	**.58 [.51-.66]***
Hispanic (*n* = 100)	**.55 [.40-.70]***	.40 [.22-.56]	.54 [.37-.72]	.47 [.34-.60]

Note. AUC = Area Under the Curve; CI = 95% Confidence Intervals; **bold** * indicates small effect size, ** denotes medium effect size, and *** denotes a large effect size.

Table 28

AUC Difference Tests for Dynamic Risk Assessment for Offender Re-Entry Scores by Gender and Race for New Offences

	Stable		Acute		Protective		DRAOR Total	
	D	*p*	*D*	*p*	*D*	*p*	*D*	*p*
Comparisons across Race								
Men								
White v. Black	0.606	.545	0.279	.781	-0.177	.859	0.404	.686
White v. Hispanic	1.270	.204	0.382	.702	-0.620	.535	0.685	.494
Black v. Hispanic	0.968	.333	.257	.797	-0.526	.599	0.496	.620
Women								
White v. Black	-1.127	.260	-1.735	.174	-0.084	.933	-1.002	.316
White v. Hispanic	-0.095	924	1.063	.288	-0.0634	.949	0.526	.599
Black v. Hispanic	0.318	.751	1.551	.121	-0.030	.976	0.878	.380
Comparisons across Gender								
White – Men v Women	1.084	.278	1.165	.244	0.073	.942	1.013	.311
Black – Men v. Women	-0.862	.389	-0.792	.429	0.090	.928	-0.647	.518
Hispanic – Men v. Women	-0.651	.515	0.787	.432	0.349	.727	0.170	.865

Note. D = coefficient for difference test, *p* represents probability of results using a two-tailed significance test.

Any return. As illustrated in Table 29, DRAOR subscale and Total scores modestly predicted returns for most gender by race groups, with the exception of Black men. While relative predictive accuracy for DRAOR scores was virtually identical for White men and women, slight (i.e., non-significant) differences were observed for Black and Hispanic men and women (see Table 30). The predictive ability of Stable, Acute, and DRAOR Total scores was marginally higher for Black women when compared to Black men, though Protective scores predicted weakly for both genders. Consistent with results for the technical violations outcome, discrimination was highest for Stable scores (and correspondingly, DRAOR Total scores) for Hispanic JIPs. Again here, it is important to keep the confidence intervals in mind when considering the magnitude of the *AUC*s. Though several of the *AUC*s for Hispanic men and women represent medium and large effects, the associated confidence intervals are too wide to be able to confidently conclude that the observed *AUC*s accurately represent the true effect.

Caveats aside, a similar pattern of relative predictive accuracy was apparent for Hispanic men and women in that Stable scores were most predictive, and Acute scores provided the least discrimination. Finally, results of difference tests revealed that discrimination was not impacted by gender but did show some variability according to race. Interestingly, discrimination was similar for all women, regardless of race, but this was not the case for men. Acute and DRAOR Total scores more accurately discriminated recidivists from non-recidivists for White men when compared to Black men, with $D_{\text{Acute}} = 2.50$, $p = .012$ and $D_{\text{Total}} = 2.55$, $p = .011$. As with the technical violation outcome, Stable scores predicted returns more accurately for Hispanic men in comparison with their White ($D = -2.18$, $p = .029$) and Black ($D = -2.81$, $p = .005$) counterparts. DRAOR

Total scores were also better at differentiating recidivists and non-recidivists for Hispanic men as compared to Black men (D = -2.31, p = .021).

Table 29

Discriminative Validity of Dynamic Risk Assessment for Offender Re-Entry Scores by Gender and Race for Any Return

	Stable *AUC* [CI]	Acute AUC [CI]	Protective AUC [CI]	DRAOR Total AUC [CI]
Men				
White (*n* = 2,194)	**.59 [.57-.62]***	**.59 [.57-.62]***	**.58 [.55-.60]***	**.61 [.58-.63]***
Black (*n* = 491)	.54 [.48-.59]	.52 [.46-.57]	.52 [.46-.57]	.53 [.48-.59]
Hispanic (*n* = 78)	**.72 [.60-.84]*****	**.59 [.44-.73]***	**.63 [.51-.76]***	**.69 [.55-.82]****
Women				
White (*n* = 2,167)	**.59 [.57-.62]***	**.60 [.57-.62]***	**.58 [.55-.60]***	**.61 [.58-.63]***
Black (*n* = 496)	**.55 [.49-.61]***	**.55 [.49-.61]***	.52 [.46-.58]	**.55 [.49-.60]***
Hispanic (*n* = 100)	**.65 [.53-.78]****	.54 [.41-.67]	**.55 [.42-.68]***	**.61 [.49-.74]***

Note. AUC = Area Under the Curve; CI = 95% Confidence Intervals; **bold** * indicates small effect size, ** denotes medium effect size, and *** denotes a large effect size.

Table 30

AUC Difference Tests for Dynamic Risk Assessment for Offender Re-Entry Scores by Gender and Race for Any Return

	Stable		Acute		Protective		DRAOR Total	
	D	*p*	*D*	*p*	*D*	*p*	*D*	*p*
Comparisons across Race								
Men								
White v. Black	1.711	.087	**2.500**	**.012**	1.897	.058	**2.547**	**.011**
White v. Hispanic	**-2.177**	**.029**	0.087	.931	-0.889	.374	-1.260	.208
Black v. Hispanic	**-2.811**	**.005**	-0.944	.345	-1.642	.101	**-2.305**	**.021**
Women								
White v. Black	1.245	.213	1.584	.113	1.745	0.081	1.893	.058
White v. Hispanic	-0.982	.326	0.971	.332	0.495	.621	-0.083	.934
Black v. Hispanic	-1.518	.129	0.146	.884	-0.381	.703	-1.001	.317
Comparisons across Gender								
White – Men v Women	-0.226	.821	-0.396	.692	-0.168	.867	0.000	1.00
Black – Men v. Women	-0.374	.708	-0.769	.442	-0.074	.941	-0.373	.710
Hispanic – Men v. Women	0.848	.396	0.542	.600	0.972	.331	0.870	.385

Note. D = coefficient for difference test, *p* represents probability of results using a two-tailed significance test and **bold** denotes a significant different test.

Calibration

The absolute predictive accuracy of the DRAOR subscale and Total scores was examined using the E/O index and 95% confidence intervals. Calibration was assessed separately for technical violations, new offences, and any return for each of the six race by gender study groups. As described earlier, values for the numerator of the E/O index (i.e., the number of recidivists expected) were derived using base rates observed for 343 White men included in earlier evaluation of the DRAOR in Iowa (Chadwick, 2014). This dataset was selected as a comparison dataset for two main reasons. First, this data was collected from parole officers who had recently received training on how to score the DRAOR by the scale's developer; thus, confidence in scoring fidelity is high. In the absence of empirically validated and published norms, this was the next best option. As shown in Table 31 below, DRAOR subscale and Total scores distinguished between recidivists and non-recidivists with greater accuracy in Chadwick's (2014) pilot sample compared to the discrimination observed using the current sample. Second, using the pilot dataset as a reference point allowed for a more nuanced understanding of how gender and race impact calibration in this sample as it provides a baseline for level of predictive accuracy that should be expected for the male contingent of the current sample. For example, finding that calibration is poor for Black women but also for White men has different implications in terms of the causes of poor absolute predictive ability than does finding poor calibration for Black women but excellent calibration for White men.

Table 31

Relative Predictive Accuracy of the DRAOR Subscale and Total Scores by Type of Recidivism

	Technical Violation *AUC* [95% CI]	New Offence *AUC* [95% CI]	Any Return *AUC* [95% CI]
Total	.736 [.683 – .789]	.588 [.525 – .650]	.742 [.688 – .796]
Stable	.686 [.630 – .742]	.557 [.493 – .622]	.694 [.637 – .751]
Acute	.713 [.659 – .767]	.614 [.552 – .676]	.735 [.681 – .790]
Protective	.734 [.680 – .788]	.582 [.518 – .645]	.728 [.671 – .784]

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry, *AUC* = Area Under the Curve statistic, CI = confidence interval.

Initial Calibration Analyses. At the outset, this study attempted to evaluate the absolute predictive accuracy of all possible DRAOR subscale and Total scores (i.e., 0-12 for the Stable and Protective subscale, 0-14 for the Acute subscale, and -12 to 26 for the DRAOR Total scores) for all three recidivism outcomes broken down by gender and race. While these analyses constituted an interesting academic exercise, inadequate sample size for Black and Hispanic JIPs produced results of little practical significance. Accordingly, the initial approach (examining the calibration for each possible score) had to be revised.

In principle, statistical significance is determined through the E/O index confidence interval (CI), such that CIs containing 1 indicate that expected and observed recidivism rates are not significantly different (i.e., good calibration; Hanson, 2017; Helmus & Babchishin, 2017). However, this can become problematic when sample size is low. The width of CIs is strongly affected by sample size and variability in the sample (Hazra, 2017; Tabachnick & Fidell, 2013), thus smaller samples tend to produce much wider CIs. In the context of calibration analyses, this means that E/O indexes calculated using smaller samples often have very broad CIs and are therefore significant, yet

imprecise. Extreme examples of this were common for Hispanic JIPs. For instance, for Hispanic men (*n* = 78), the E/O index associated with a score of 1 on the Stable domain was 1.03. At first glance, this would appear to indicate virtually perfect calibration. However, the 95% confidence interval associated with this statistic was 0.15 (lower bound) to 7.33 (upper bound). Clearly, this E/O cannot be considered to be reliable. Low sample size was also problematic as it was simply not possible to assess calibration for each possible score. For Black and Hispanic women, not all subscale and Total scores were observed in the sample, and this, coupled with low base rates for the new offence outcome in particular precluded the calculation of multiple E/O indexes. Results of these initial calibration analyses are included in Appendix H for the consideration of interested readers.

Revised Calibration Analyses. In view of the foregoing, DRAOR subscale and Total scores were collapsed into bins to increase stability and precision of results. Stable, Acute, and Protective scores were recoded into three groups (low, mean, and high), and the four risk bins employed in Study 2 (i.e., low-moderate, moderate, moderate-high, and high) were used to categorize DRAOR Total scores. The rationale for using low, mean, and high rather than low, moderate, and high as categories for the subscales was to attempt to more evenly distribute the number of individuals in each group. Individuals in the mean group were those who had scored within 0.5 standard deviation (SD) of the mean, and those in the low or high groups had scored more than 0.5 SD away from the mean. Overall calibration (i.e., a cumulative E/O index) for each race by gender subsample was also assessed. Results are presented by subsample.

White men. Contrary to expectations, results indicated that calibration was relatively poor. The absolute predictive accuracy of DRAOR subscale and Total scores was inconsistent across subscale scores, with many significant E/O indexes (see Table 32). Subscale scores did not provide consistent levels of absolute predictive accuracy across the recidivism outcomes. For example, high Stable scores demonstrated good calibration when predicting technical violations (E/O = 1.05), but not when predicting new offences (E/O = 2.42). Significance (or non-significance) aside, one pattern did emerge for the technical violations and any return outcomes; the E/O indexes for the Stable and Acute subscale scores demonstrate that recidivists tended to be underestimated at lower scores (i.e., E/O index less than 1 indicates that there were more recidivists than predicted) and overestimated at higher scores (i.e., E/O index greater than 1). The opposite trend (i.e., recidivism was underestimated for those with high scores) was observed for Protective scores. The CIs associated with the E/O indexes assessing calibration of the Stable, Acute, and Protective subscales for both the technical violations and any return outcomes are quite narrow, suggesting that results for these two types of recidivism are reliable. Collectively, findings suggest that in the current sample, men rated as having lower levels of risk and higher levels of protective factors incurred more technical violations and returned to custody for any reason more than expected and that the opposite was true for those rated as high on the Stable and Acute risk domains and as low on the Protective domain.

Calibration was especially poor for the new offence outcome, with only one non-significant E/O index (high DRAOR Total scores, E/O = 3.50). An E/O index of 3.50 indicates that the expected number of recidivists was 3.5 times higher than the actual

number of recidivists. Notably, however, the CIs associated with this E/O statistic were too broad to allow confident interpretation of these results. Otherwise, all E/O index statistics were significant, indicating that the number of expected and observed recidivists differed significantly. For this outcome, DRAOR subscale and Total scores overestimated the number of JIPs who would commit new offences, frequently predicting two and three times the actual rate of new offences.

Table 32

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry Subscale and Total Scores – White Men (n = 2,194)

	Technical Violation			New Offence			Any Return		
		95% CI			95% CI			95% CI	
	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound
Stable									
Low	0.82	0.72	0.93	2.69	2.04	3.56	**0.93**	**0.82**	**1.05**
Mean	0.61	0.52	0.72	2.26	1.56	3.28	0.80	0.68	0.94
High	**1.05**	**0.99**	**1.12**	2.42	2.11	2.77	1.09	1.02	1.16
Acute									
Low	0.72	0.64	0.81	1.94	1.50	2.51	0.78	0.70	0.88
Mean	**1.02**	**0.86**	**1.20**	3.10	2.12	4.52	**1.14**	**0.97**	**1.35**
High	**1.10**	**1.03**	**1.17**	2.85	2.48	3.28	1.18	1.11	1.25
Protective									
Low	1.14	1.07	1.21	2.60	2.26	2.99	1.16	1.09	1.24
Mean	**0.85**	**0.72**	**1.01**	2.22	1.49	3.32	**0.96**	**0.81**	**1.13**
High	0.56	0.50	0.63	2.25	1.74	2.92	0.69	0.62	0.78
Total									
Low-Moderate	0.76	0.68	0.86	2.50	1.91	3.28	0.86	0.77	0.97
Moderate	**1.07**	**0.98**	**1.17**	3.10	2.55	3.77	1.17	1.08	1.27
Moderate-High	1.20	1.10	1.30	2.14	1.79	2.55	1.18	1.09	1.28
High	**1.21**	**0.67**	**2.18**	**3.50**	**0.88**	**13.99**	**1.27**	**0.70**	**2.30**

Note. **Bold** denotes non-significant E/O index.

Black men. The absolute predictive accuracy of DRAOR subscale and Total scores predicting technical violations, new offences, and returns was also inconsistent for Black men (see Table 33). While more of the E/O index statistics were non-significant when compared to White men, the degree of variability seen (i.e., the extent of under- and over-prediction) for the technical violations and any return outcomes was likewise more pronounced. For example, for the any return outcome, an E/O index of 0.27 was calculated for low Protective scores, meaning that this range of scores only predicted 27% of the observed recidivists. That said, low and mean Stable and Acute scores and high protective scores underestimated technical violations and returns, and high risk/low protective factor scores overpredicted recidivism, which is consistent with the results for White men.

Across all three outcomes, Stable scores appeared to demonstrate the best absolute predictive accuracy, though in many cases the ratio of expected to observed recidivists was quite far from 1.0. Unsurprisingly, calibration was weakest for the prediction of new offences. Despite several non-significance E/O indexes, calibration could best be described as unreliable for DRAOR scores predicting this outcome. CIs were imprecise (i.e., wider than 1.0) and E/O ratios suggested that recidivists were over-estimated considerably. The two exceptions to this were mean Protective scores, with E/O = 1.11 and high DRAOR Total scores, with E/O = 0.67.

Table 33

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry Subscale and Total Scores – Black Men (*n* = 491)

	Technical Violation			New Offence			Any Return		
		95% CI			95% CI			95% CI	
	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound
Stable									
Low	0.73	0.54	0.98	**1.58**	**0.92**	**2.73**	**0.80**	**0.60**	**1.07**
Mean	0.69	0.48	0.99	**1.50**	**0.81**	**2.79**	**0.89**	**0.62**	**1.27**
High	**1.09**	**0.96**	**1.24**	1.89	1.48	2.42	**1.09**	**0.97**	**1.24**
Acute									
Low	0.64	0.50	0.82	**1.59**	**0.92**	**2.74**	0.70	0.54	0.90
Mean	**0.90**	**0.66**	**1.21**	2.05	1.14	3.71	**1.01**	**0.75**	**1.36**
High	1.19	1.03	1.36	2.06	1.61	2.64	1.21	1.06	1.38
Protective									
Low	1.19	1.04	1.37	1.90	1.49	2.44	0.27	0.23	0.30
Mean	**0.98**	**0.67**	**1.46**	**1.11**	**0.60**	**2.06**	**1.16**	**0.80**	**1.68**
High	0.47	0.37	0.60	2.02	1.17	3.47	2.03	1.61	2.58
Total									
Low-Moderate	0.63	0.49	0.82	**1.43**	**0.87**	**2.33**	0.72	0.56	0.94
Moderate	**1.08**	**0.91**	**1.27**	2.48	1.77	3.47	**1.13**	**0.96**	**1.32**
Moderate-High	1.38	1.13	1.68	1.68	1.19	2.36	1.31	1.09	1.59
High	**0.95**	**0.36**	**2.53**	**0.67**	**0.22**	**2.07**	**1.00**	**0.38**	**2.66**

Note. **Bold** denotes non-significant E/O index.

Hispanic men. Even more caution is required when interpreting the calibration results for Hispanic men. As discussed earlier, E/O statistic and more specifically, its CIs, are highly sensitive to small sample size. As there were only 78 Hispanic men, many of the CIs observed for these analyses were imprecise. While most of the E/O indexes for DRAOR subscale and Total scores were theoretically indicative of good calibration, results should be viewed as largely exploratory rather than confirmatory. Small sample size and/or low base rates prevented calculation of the E/O index in several cases (see Table 34).

With respect to the prediction of technical violations, calibration was reasonable overall, with over- and under-predictions tending to be by relatively small proportions. The impact of small sample size is apparent in these analyses, as the CIs for the E/O indexes calculated using the mean subscale scores were consistently the widest. Results for the new offence outcome cannot be confidently reported as CIs are highly imprecise. The only pattern discernable for this outcome was that DRAOR subscale and Total scores tended to over-predict reoffences. That said, DRAOR subscale and Total scores evidenced good calibration in the prediction of returns. While the CIs associated with the E/O statistics for the mean score groups for each subscale were wider than desired, the E/O indexes themselves were quite narrowly clustered around 1.0 with only two exceptions. Both low Acute and high Protective scores underpredicted recidivists, predicting 67% and 67% of the observed recidivists, respectively.

Table 34

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry Subscale and Total Scores – Hispanic Men (*n* = 78)

	Technical Violation			New Offence			Any Return		
		95% CI			95% CI			95% CI	
	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound
Stable									
Low	**0.87**	**0.44**	**1.75**	**2.30**	**0.58**	**9.21**	**1.01**	**0.50**	**2.01**
Mean	**0.91**	**0.34**	**2.42**	**0.91**	**0.29**	**2.82**	**0.97**	**0.40**	**2.33**
High	**0.89**	**0.65**	**1.21**	2.81	1.26	6.26	**0.96**	**0.70**	**1.31**
Acute									
Low	0.61	0.38	0.98	**1.11**	**0.46**	**2.66**	**0.67**	**0.42**	**1.08**
Mean	**0.88**	**0.33**	**2.35**	**--**	**--**	**--**	**1.01**	**0.38**	**2.70**
High	**1.05**	**0.74**	**1.50**	2.94	1.32	6.54	**1.14**	**0.81**	**1.61**
Protective									
Low	**1.04**	**0.76**	**1.42**	2.16	1.13	4.16	**1.08**	**0.79**	**1.47**
Mean	**0.92**	**0.38**	**2.22**	**1.04**	**0.26**	**4.16**	**1.06**	**0.44**	**2.55**
High	**0.54**	**0.27**	**1.08**	**--**	**--**	**--**	**0.68**	**0.34**	**1.36**
Total									
Low-Moderate	**0.81**	**0.44**	**1.50**	**2.56**	**0.64**	**10.25**	**0.94**	**0.51**	**1.75**
Moderate	**0.89**	**0.61**	**1.29**	**2.21**	**0.99**	**4.91**	**0.99**	**0.68**	**1.44**
Moderate-High	**1.15**	**0.70**	**1.91**	**2.33**	**0.75**	**7.23**	**1.13**	**0.69**	**1.85**
High	**--**	**--**	**--**	**--**	**--**	**--**	**--**	**--**	**--**

Note. **Bold** denotes non-significant E/O index, -- indicates that E/O could not be calculated due to low sample size or base rate.

White women. The absolute predictive accuracy of DRAOR subscale and Total scores predicting recidivism was poor overall, with the majority of E/O indexes indicating that the number of expected and observed recidivists were significantly different (see Table 35). Notably, sample size was sufficient for these analyses and that results regarding calibration (or lack of calibration) for the technical violations and any return outcome are reliable. Unfortunately, low base rates for the new offence outcome were still a factor and likely negatively impacted precision for those analyses.

Interestingly, calibration appeared to be better for higher risk White women. For technical violations, absolute predictive accuracy was better for women with high Stable and Acute scores (E/O index = 0.99 and 1.04, respectively), and low Protective scores (E/O index = 1.05). For lower risk women (i.e., those with low Stable and Acute scores and high Protective scores), technical violations were underpredicted, sometimes substantially (e.g., high protective scores predicted only 56% of observed recidivists). A similar pattern was observed for the any return outcome. Also, though calibration was inconsistent for DRAOR Total scores predicting technical violations and returns, a similar progression from underprediction at lower levels of risk to overprediction at highest levels of risk was observed.

With respect to new offences, neither DRAOR subscale nor Total scores demonstrated good absolute predictive validity. All E/O indexes indicated that reoffences were overpredicted.

Table 35

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry Subscale and Total Scores – White Women (*n* = 2,167)

	Technical Violation			New Offence			Any Return		
		95% CI			95% CI			95% CI	
	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound
Stable									
Low	0.80	0.71	0.91	3.29	2.39	4.52	**0.90**	**0.80**	**1.03**
Mean	0.55	0.46	0.65	2.44	1.61	3.71	0.71	0.60	0.84
High	**0.99**	**0.93**	**1.05**	2.82	2.44	3.27	**1.04**	**0.98**	**1.11**
Acute									
Low	0.72	0.64	0.82	2.05	1.52	2.76	0.77	0.67	0.87
Mean	**0.95**	**0.80**	**1.13**	3.92	2.47	6.23	**1.07**	**0.90**	**1.28**
High	**1.04**	**0.98**	**1.10**	3.43	2.96	3.97	1.13	1.07	1.20
Protective									
Low	**1.05**	**0.99**	**1.12**	3.06	2.61	3.58	1.09	1.03	1.16
Mean	0.75	0.64	0.87	2.15	1.45	3.19	0.85	0.73	0.99
High	0.56	0.50	0.62	2.65	2.03	3.45	0.69	0.62	0.77
Total									
Low-Moderate	0.69	0.61	0.77	2.28	1.75	2.96	0.78	0.69	0.87
Moderate	**1.02**	**0.93**	**1.11**	3.99	3.17	5.03	1.11	1.02	1.21
Moderate-High	1.13	1.05	1.23	2.58	2.14	3.12	1.16	1.07	1.25
High	**1.53**	**0.89**	**2.64**	10.50	1.48	74.54	**1.62**	**0.94**	**2.78**

Note. **Bold** denotes non-significant E/O index.

Black women. For Black women, calibration was best for higher risk women (see Table 36). As for White women, absolute predictive accuracy when predicting technical violations and any return was highest for Black women with high Stable and Acute, and low Protective scores, with E/O indexes in the range of 0.93 to 1.06. Conversely, technical violations and returns were consistently underestimated for low risk women, with E/O indexes ranging from 0.50 to 0.74 for women with Stable and Acute scores at or below the mean, and from 0.44 to .74 for women with average or higher Protective scores. An examination of the calibration of DRAOR Total scores predicting technical violations and any returns supported this conclusion, with moderate and moderate-high scores yielding the best absolute predictive accuracy, as evidenced by E/O indexes grouped most tightly around 1.0.

With respect to new offences, all of the DRAOR scores examined were poorly calibrated. Recidivists were over-predicted by factors ranging between 2.54 and 5.66.

Table 36

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry Subscale and Total Scores – Black Women (*n* = 496)

	Technical Violation			New Offence			Any Return		
		95% CI			95% CI			95% CI	
	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound
Stable									
Low	0.58	0.45	0.75	4.44	1.85	10.68	0.66	0.51	0.85
Mean	0.50	0.37	0.68	2.55	1.14	5.67	0.63	0.47	0.85
High	**0.93**	**0.82**	**1.04**	2.89	2.13	3.93	**0.98**	**0.87**	**1.10**
Acute									
Low	0.55	0.44	0.70	5.02	1.88	13.38	0.61	0.48	0.77
Mean	0.74	0.57	0.97	3.00	1.50	6.00	**0.84**	**0.64**	**1.09**
High	**0.97**	**0.86**	**1.10**	3.23	2.37	4.41	**1.06**	**0.93**	**1.20**
Protective									
Low	**0.98**	**0.86**	**1.11**	3.02	2.21	4.14	**1.03**	**0.91**	**1.16**
Mean	0.66	0.48	0.91	5.66	1.42	22.63	0.74	0.54	1.01
High	0.44	0.35	0.55	2.61	1.45	4.72	0.54	0.44	0.67
Total									
Low-Moderate	0.51	0.41	0.64	4.23	1.90	9.42	0.59	0.47	0.74
Moderate	**0.93**	**0.80**	**1.09**	3.34	2.24	4.98	**1.01**	**0.87**	**1.18**
Moderate-High	**1.08**	**0.91**	**1.29**	2.54	1.67	3.86	**1.11**	**0.93**	**1.31**
High	**1.11**	**0.50**	**2.47**	**--**	**--**	**--**	**1.17**	**0.52**	**2.60**

Note. **Bold** denotes non-significant E/O index, -- indicates that E/O could not be calculated due to low sample size or base rate.

Hispanic women. Results of the calibration analyses for Hispanic women are similar to those described above for Hispanic men. As illustrated in Table 37, the majority of the E/O indexes calculated for DRAOR subscale and Total scores predicting the three recidivism outcomes were non-significant and therefore indicative, in principle, of good calibration. However, the breadth of many of the CIs raises concerns about precision and reliability. As with White and Black women, calibration appears to be best (i.e., E/O index close to 1.0 coupled with sufficiently narrow CIs) for higher Stable and Acute risk scores and low Protective scores when predicting technical violations and any returns. Equally, moderate and moderate-high DRAOR Total scores are associated with the best absolute predictive accuracy for these outcomes.

The combination of low sample size and low base rates for new offences largely precluded analyses for this outcome, though the E/O indexes that could be calculated suggest that new offences were typically overestimated for Hispanic women.

Table 37

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry Subscale and Total Scores – Hispanic Women (*n* = 100)

	Technical Violation			New Offence			Any Return		
		95% CI			95% CI			95% CI	
	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound	E/O Index	Lower Bound	Upper Bound
Stable									
Low	**0.92**	**0.46**	**1.85**	--	--	--	**1.06**	**0.53**	**2.13**
Mean	**0.95**	**0.39**	**2.27**	--	--	--	**1.26**	**0.52**	**3.03**
High	**0.92**	**0.70**	**1.21**	3.81	1.71	8.47	**0.96**	**0.73**	**1.25**
Acute									
Low	**0.77**	**0.43**	**1.39**	**4.51**	**0.64**	**32.02**	**0.78**	**0.44**	**1.37**
Mean	**0.78**	**0.35**	**1.75**	--	--	--	**0.90**	**0.40**	**2.01**
High	**1.06**	**0.80**	**1.40**	5.49	2.28	13.19	**1.16**	**0.87**	**1.53**
Protective									
Low	**1.06**	**0.80**	**1.42**	4.81	2.00	11.57	**1.11**	**0.84**	**1.48**
Mean	**0.94**	**0.42**	**2.09**	**2.54**	**0.36**	**18.04**	**0.93**	**0.44**	**1.94**
High	0.54	0.31	0.95	--	--	--	**0.68**	**0.39**	**1.20**
Total									
Low-Moderate	**0.65**	**0.38**	**1.12**	--	--	--	**0.76**	**0.44**	**1.31**
Moderate	**1.10**	**0.74**	**1.62**	3.02	1.26	7.27	**1.13**	**0.78**	**1.65**
Moderate-High	**1.10**	**0.75**	**1.60**	11.99	1.69	85.11	**1.15**	**0.79**	**1.68**
High	--	--	--	--	--	--	--	--	--

Note. **Bold** denotes non-significant E/O index, -- indicates that E/O could not be calculated due to low sample size or base rate.

Summary calibration analyses. To increase the stability and interpretability of results, a series of overarching calibration analyses assessing the absolute predictive accuracy of DRAOR scores predicting each of the three recidivism outcomes were conducted for each gender by race subgroup. Results are presented in Table 38. Considered collectively, DRAOR scores tended to underestimate technical violations and any returns and to overestimate new offences, regardless of JIPs' race and gender. With respect to technical violations, calibration was reasonable for Black and Hispanic men, and for Hispanic women, with DRAOR scores predicting 91%, 84% and 86% of recidivists, respectively. Interestingly, though the proportion of expected recidivists was derived from a sample of White men, DRAOR scores did not demonstrate strong absolute predictive validity for the White men in this sample. Technical violations were underpredicted even more severely for White and Black women. With respect to new offences, calibration was poor for all gender by race subgroups, and new offences expected were overpredicted in all cases. New offences were overestimated to the greatest extent for Hispanic women (E/O index = 5.05) and least for Black men (E/O index = 1.73). DRAOR scores predicting any return demonstrated the best overall calibration across all six race by gender groups. The number of expected and observed recidivists was not significantly different for White men (E/O index = 0.97), Black men (E/O index = 0.95), Hispanic men (E/O index = 0.91), and Hispanic women (E/O index = 0.92), and while the number of expected and observed recidivists did differ for White and Black women, the obtained E/O indexes did not suggest gross deviations from expectations.

Table 38

Absolute Predictive Accuracy of Dynamic Risk Assessment for Offender Re-Entry (DRAOR) Total Scores by Race and Gender

Recidivism Outcome	Sample	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Technical Violations	White Men	0.62	0.56	1358.00	1222.06	0.90	0.85	0.95
	Black Men	0.62	0.56	302.00	273.49	**0.91**	**0.81**	**1.01**
	Hispanic Men	0.67	0.56	52.00	43.45	**0.84**	**0.64**	**1.10**
	White Women	0.66	0.56	1427.00	1207.02	0.85	0.80	0.89
	Black Women	0.74	0.56	368.00	276.27	0.75	0.68	0.83
	Hispanic Women	0.65	0.56	65.00	55.70	**0.86**	**0.67**	**1.09**
New Offence	White Men	0.13	0.30	279.00	664.78	2.38	2.12	2.68
	Black Men	0.18	0.30	86.00	148.77	**1.73**	**1.40**	**2.14**
	Hispanic Men	0.14	0.30	11.00	23.63	2.15	1.19	3.88
	White Women	0.11	0.30	237.00	656.60	2.77	2.44	3.15
	Black Women	0.10	0.30	52.00	150.29	2.89	2.20	3.79
	Hispanic Women	0.06	0.30	6.00	30.30	5.05	2.27	11.24
Any Return	White Men	0.64	0.62	1408.00	1362.47	**0.97**	**0.92**	**1.02**
	Black Men	0.66	0.62	322.00	304.91	**0.95**	**0.85**	**1.06**
	Hispanic Men	0.68	0.62	53.00	48.44	**0.91**	**0.70**	**1.20**
	White Women	0.68	0.62	1464.00	1345.71	0.92	0.87	0.97
	Black Women	0.76	0.62	376.00	308.02	0.82	0.74	0.91
	Hispanic Women	0.68	0.62	67.00	61.48	**0.92**	**0.72**	**1.17**

Note. **Bold** denotes non-significant E/O index.

Logistic Regression

A series of direct logistic regression analyses were performed separately for the technical violations, new offences, and any return outcomes. Importantly, these analyses differ from earlier discrimination and calibration analyses which also fall under the broader umbrella of prediction analyses. Discrimination and calibration analyses assess relative and absolute predictive accuracy at the level of the individual (i.e., is one individual more likely to recidivate than another, how many recidivists are expected in total), whereas regression analyses instead focus on how assessment scores are related to outcome, and the extent to which changes in scores impacts likelihood of recidivism.

Prior to examining the predictive abilities of individual items, tests of the full (omnibus) model containing Stable, Acute, and Protective factor subscale scores (entered as a set) were conducted. Results of these omnibus models are shown in Table 39 below. The full model was statistically reliable (i.e., significant) for White men for all three recidivism outcomes, with χ^2 (3, N = 2,194) = 51.92, $p < .001$, χ^2 (3, N = 2,194) = 15.63, $p = .001$, and χ^2 (3, N = 2,194) = 69.09, $p < .001$, for technical violations, new offences, and any return, respectively. The set of Stable, Acute, and Protective factor subscale scores also predicted technical violations (χ^2 (3, N = 2,167) = 70.55, $p < .001$) and any returns ^{2} (3, N = 2,167) = 71.13, $p < .001$) for White women. The model approached significance for Hispanic men when predicting technical violations and any return but did not maintain significance following the required adjustments to significance levels when using multiple tests. The set of DRAOR subscale scores did not reliably distinguish recidivists from non-recidivists for Black JIPs or Hispanic women, regardless of outcome, and the model's ability to predict new offences was generally poor.

Importantly, even in the significant models, the set of DRAOR subscales scores did not account for much of the variance in recidivism outcome as none of the McFadden's ρ^2 values approached 0.2 (i.e., the benchmark for a small effect). Overall, these results suggest that the set of DRAOR subscale scores contribute little to the prediction of recidivistic outcomes for JIPs in the sample. In other words, DRAOR subscale scores did not play a large role in the prediction of the recidivism outcomes examined in this study.

Table 39

Omnibus Logistic Regression Analysis Predicting Outcome from Set of Stable, Acute, and Protective Subscale Scores

			-2 Log Likelihood	χ^2	*df*	Sig.	McFadden's ρ^2
Technical Violations	Men	White (*n* = 2,194)	2864.22	51.92	3	**< .001***	.023
		Black (*n* = 491)	653.32	1.11	3	.781	< .001
		Hispanic (*n* = 78)	86.91	12.39	3	**.006**	.072
	Women	White (*n* =2,167)	2711.95	70.55	3	**< .001***	.031
		Black (*n* = 496)	562.81	3.65	3	.302	.006
		Hispanic (*n* =100)	125.53	3.96	3	.266	.014
New Offence	Men	White (*n* = 2,194)	1656.03	15.63	3	**.001***	.005
		Black (*n* = 491)	452.88	2.73	3	.435	.005
		Hispanic (*n* = 78)	60.47	3.0	3	.392	< .001
	Women	White (*n* =2,167)	1490.28	5.78	3	.123	.001
		Black (*n* = 496)	326.56	6.43	3	.096	.008
		Hispanic (*n* =100)	43.12	2.27	3	.519	< .001
Any Return	Men	White (*n* = 2,194)	2793.66	69.09	3	**< .001***	.031
		Black (*n* = 491)	630.24	1.96	3	.582	.002
		Hispanic (*n* = 78)	86.84	11.01	3	**.012**	.078
	Women	White (*n* =2,167)	2659.94	71.13	3	**< .001***	.032
		Black (*n* = 496)	543.94	4.93	3	.177	.007
		Hispanic (*n* =100)	119.41	5.19	3	.159	.039

Note. **Bold** denotes significance at the alpha = .05 level; **bold*** denotes significance following correction to *p* values to account for multiple corrections.

Prediction at the subscale level. Tables 40, 41, and 42 show regression coefficients, Wald statistics, odds ratios, and pseudo R^2 (Mcfadden's ρ^2) values for each of the DRAOR subscales. Results are disaggregated by study group and outcome.

Technical violations. As illustrated in Table 40, according to the Wald criterion, Stable, Acute, and Protective scores reliably predicted technical violations for White men ($z_{\text{Stable}} = 33.70$, $p < .001$, $z_{\text{Acute}} = 39.70$, $p < .001$, and $z_{\text{Protective}} = 28.19$, $p < .001$) and White women ($z_{\text{Stable}} = 49.13$, $p < .001$, $z_{\text{Acute}} = 59.22$, $p < .001$, and $z_{\text{Protective}} = 35.49$, $p < .001$). Stable scores also reliably predicted technical violations for Hispanic men ($z = 9.05$, $p = .003$) but not Hispanic women ($z = 3.19$, $p = .074$). Individual DRAOR subscale scores did not predict technical violations for Black JIPs. Notably, consideration of the odds ratios and McFadden's ρ^2 values associated with the predictors that did achieve significance again suggest that effects were quite small. For example, for White men, the odds ratio for Stable scores is 1.07. This indicates that the change in likelihood of recidivism associated with a one-point change in Stable scores is minimal. When coupled with a McFadden's ρ^2 of .016, it is clear that Stable scores play a small role in the prediction of technical violations. This logic applies to all of the models predicting technical violations for White men and women. However, for Hispanic men, Stable scores predict technical violations more strongly, with an odds ratio (*OR*) of 1.42 and McFadden's $\rho^2 = .132$. While this rho value is still below the .2 cut-off typically used to indicate a modest effect size, it approaches the cut-off and is only one to exceed .1. Given earlier findings regarding the relative predictive accuracy (i.e., discrimination; $AUC = .73$) of Stable scores for Hispanic men, these results are unsurprising.

Table 40

Logistic Regression Analysis of Technical Violations as a Function of Dynamic Risk Assessment for Offender Re-Entry Subscale Scores

			B	S.E.	Wald test (*z*-ratio)	*df*	Sig.	Odds Ratio	95% CI	McFadden's ρ^2
	White	Stable	.096	.017	33.699	1	**< .001**	1.101	1.066 – 1.137	.016
		Acute	.099	.016	39.703	1	**< .001**	1.104	1.104 – 1.071	.018
		Protective	-.083	.016	28.186	1	**< .001**	.921	.893 – .949	.013
Men	Black	Stable	.028	.036	.620	1	.431	1.029	.959 – 1.104	.001
		Acute	-.006	.036	.031	1	.859	.994	.926 – 1.067	< .001
		Protective	-.016	.034	.226	1	.634	.984	.920 – 1.052	< .001
	Hispanic	Stable	.349	.116	9.049	1	**.003**	1.418	1.129 – 1.779	.132
		Acute	.067	.082	.658	1	.417	1.069	.910 – 1.255	.008
		Protective	-.135	.107	1.603	1	.205	.873	.708 – 1.077	.021
	White	Stable	.119	.017	49.134	1	**< .001**	1.126	1.089 – 1.164	.023
		Acute	.123	.016	59.221	1	**< .001**	1.130	1.096 – 1.166	.028
		Protective	-.096	.016	35.490	1	**< .001**	.909	.880 – .938	.017
Women	Black	Stable	.064	.038	2.923	1	.087	1.066	.991 – 1.148	.006
		Acute	.063	.038	2.786	1	.095	1.065	.989 – 1.146	.006
		Protective	-.036	.038	.891	1	.345	.965	.895 – 1.040	.002
	Hispanic	Stable	.142	.079	3.194	1	.074	1.152	.986 – 1.346	.033
		Acute	.056	.071	.617	1	.432	1.057	.920 – 1.215	.006
		Protective	-.026	.075	.125	1	.724	.974	.841 – 1.127	.001

Note. CI = confidence interval, **bold** denotes significance after adjusting *p* values to control for multiple tests.

New offences and any return. DRAOR subscale scores did not predict new offences, with one exception. As shown in Table 41 below, only Stable scores for White men reliably predicted this type of recidivism ($z = 15.38$, $p < .001$), and again, consideration of the odds ratio (1.10) and pseudo R^2 (.007) indicate that a one unit change in Stable score had little impact on likelihood of reoffending. Results of the regression analyses predicting any return from DRAOR subscale scores mirror those described above for technical violations (see Tables 40 and 42), with Stable, Acute, and Protective scores significantly (albeit largely inconsequentially) predicting any return, and Stable scores predicting returns for Hispanic men.

Table 41

Logistic Regression Analysis of New Offences as a Function of Dynamic Risk Assessment for Offender Re-Entry Subscale Scores

			B	S.E.	Wald test (*z*-ratio)	*df*	Sig.	Odds Ratio	95% CI	McFadden's ρ^2
	White	Stable	.094	.024	15.375	1	**< .001**	1.099	1.048 – 1.152	.007
		Acute	.048	.022	4.699	1	.030	1.049	1.005 – 1.096	.002
		Protective	-.046	.023	4.158	1	.041	.955	.913 – .998	.002
Men	Black	Stable	.070	.046	2.334	1	.127	1.073	.980 – 1.174	.005
		Acute	.037	.046	.625	1	.429	1.037	.947 – 1.136	.001
		Protective	-.056	.044	1.600	1	.206	.946	.868 – 1.031	.003
	Hispanic	Stable	-.074	.128	.334	1	.563	.929	.723 – 1.193	.004
		Acute	.007	.110	.004	1	.949	1.007	.811 – 1.250	< .001
		Protective	-.171	.147	1.354	1	.245	.843	.632 – 1.124	.018
	White	Stable	.048	.025	3.603	1	.058	1.049	.998 – 1.102	.002
		Acute	.012	.023	.253	1	.615	1.012	.966 – 1.060	< .001
		Protective	-.046	.024	3.602	1	.058	.955	.911 – 1.002	.002
Women	Black	Stable	.124	.055	5.079	1	.024	1.132	1.016 – 1.260	.010
		Acute	.099	.053	3.492	1	.062	1.104	.995 – 1.224	.007
		Protective	-.038	.055	.496	1	.481	.962	.864 – 1.071	.001
	Hispanic	Stable	.059	.154	.148	1	.701	1.061	.785 – 1.433	.002
		Acute	-.127	.142	.792	1	.374	.881	.667 – 1.164	.008
		Protective	-.056	.150	.139	1	.709	.946	.705 – 1.268	.001

Note. CI = confidence interval, **bold** denotes significance after adjusting *p* values to control for multiple tests.

Table 42

Logistic Regression Analysis of Any Return as a Function of Dynamic Risk Assessment for Offender Re-Entry Subscale Scores

			B	S.E.	Wald test (*z*-ratio)	*df*	Sig.	Odds Ratio	95% CI	McFadden's ρ^2
Men	White	Stable	.118	.017	49.050	1	**< .001**	1.126	1.089 – 1.163	.023
		Acute	.114	.016	49.924	1	**< .001**	1.120	1.086 – 1.156	.023
		Protective	-.096	.016	36.418	1	**< .001**	.909	.881 – .937	.017
	Black	Stable	.051	.037	1.889	1	.169	1.052	.979 – 1.131	.004
		Acute	.029	.037	.597	1	.440	1.029	.957 – 1.107	.001
		Protective	-.016	.035	.221	1	.638	.984	.918 – 1.053	< .001
	Hispanic	Stable	.337	.116	8.476	1	**.004**	1.400	1.116 – 1.756	.123
		Acute	.088	.083	1.123	1	.289	1.092	.928 – 1.286	.015
		Protective	-.171	.110	2.414	1	.120	.843	.679 – 1.046	.032
Women	White	Stable	.124	.017	52.208	1	**< .001**	1.132	1.095 – 1.171	.025
		Acute	.122	.016	56.907	1	**< .001**	1.129	1.094 – 1.166	.026
		Protective	-.101	.016	38.176	1	**< .001**	.904	.876 – .933	.018
	Black	Stable	.069	.038	3.269	1	.071	1.072	.994 – 1.156	.007
		Acute	.078	.039	4.111	1	.043	1.081	1.003 – 1.166	.008
		Protective	-.031	.039	.620	1	.431	.970	.898 – 1.047	.001
	Hispanic	Stable	.170	.082	4.232	1	.040	1.185	1.008 – 1.393	.045
		Acute	.038	.073	.270	1	.603	1.038	.901 – 1.198	.003
		Protective	-.051	.078	.424	1	.515	.951	.817 – 1.107	.004

Note. CI = confidence interval, **bold** denotes significance after adjusting *p* values to control for multiple tests.

Prediction at the item level. The next set of analyses conducted in Study 3 involved assessing whether DRAOR items differentially predicted recidivism for White, Black, and Hispanic men and women. Separate models predicting technical violations, new offences, and any return were analyzed for each of the six race by gender groups and results are summarized in Table 43 below. Complete model information (regression coefficients, Wald statistics, odds ratios, etc.) is available in Appendix I.

Results indicated that only three DRAOR items (interpersonal relationships, living situation, and high expectations) reliably predicted technical violations and two items (substance abuse and interpersonal relationships) reliably predicted returns after corrections were made to *p* value thresholds to control for alpha inflation. Notably, these items only significantly predicted recidivism for White and Black men – they were not significant for women, regardless of race.

Table 43

Summary Table Describing Predictive Ability of Dynamic Risk Assessment for Offender Re-Entry Items by Gender, Race, and Outcome

	Technical Violations						New Offence						Any return					
DRAOR Item	White		Black		Hispanic		White		Black		Hispanic		White		Black		Hispanic	
	M	W	M	W	M	W	M	W	M	W	M	W	M	W	M	W	M	W
Peer Associations								✓										
Attitudes towards Authority				✓	✓											✓	✓	
Impulse Control																		
Problem Solving		✓																
Sense of Entitlement															✓			
Attachment with Others																		
Substance Abuse	✓	✓	✓										✓*	✓	✓			
Anger								✓										
Access to Victims					✓												✓	
Negative Mood																		
Employment		✓												✓				
Interpersonal Relationships	✓*		✓*		✓			✓					✓*		✓*		✓	
Living Situation	✓*		✓*										✓		✓			
Response to advice																		
Prosocial Identity		✓												✓				
High Expectations			✓*										✓	✓	✓			
Costs/Benefits																		
Social Support																		
Social Control																		

Note. M = men and W = women, ✓ denotes significance at the alpha = .05 level; ✓* denotes significance following correction to *p* values to account for multiple comparisons

Investigating promotive and protective effects. Table 44 summarizes the results of analyses investigating the potential promotive effect of items in the Protective subscale of the DRAOR on recidivism for White, Black, and Hispanic JI women. These analyses differ from the item-level analyses just described as the direction of the effect (i.e., whether likelihood of recidivism decreases) is considered as opposed to focusing solely on significant prediction. Odds ratios below 1.0 indicate that as scores on a given Protective item increase, recidivism decreases, and conversely, odds ratios above 1.0 indicate that higher scores on the item are associated with increased likelihood of recidivism. Additionally, as fewer comparisons are being made, the thresholds for significance are also different for these analyses.

Inconsistent with hypotheses, the six items comprising the DRAOR's Protective subscale did not appear to actually function in a promotive sense for the majority of women in the sample. Only one item, prosocial identity, was found to function in a promotive capacity (i.e., was associated with decreased recidivism), and this effect was only significant in relation to the technical violations and any return outcome for White women. Concerningly, results appeared to indicate that higher scores on many of the purported protective factors were associated with increased recidivism. That said, with the exception of the high expectations item for White women predicting any return, these findings did not achieve significance. Full model information for each protective factor (i.e., regression coefficients, Wald statistics, *p* values, etc.) is available in Appendix I.

Table 44

Impact of Six DRAOR Protective Factors on Recidivism for White, Black, and Hispanic Women

		Technical Violation			New Offence			Any Return		
	Protective Factor	Odds Ratio	Impact on recidivism	Sig?	Odds Ratio	Impact on recidivism	Sig?	Odds Ratio	Impact on recidivism	Sig?
White	Responsiveness to advice	.897	↓	No	.824	↓	No	.901	↓	No
	Prosocial Identity	**.816**	↓	**Yes**	.903	↓	No	**.817**	↓	**Yes**
	High Expectations	1.190	↑	No	1.270	↑	No	**1.217**	↑	**Yes**
	Costs/Benefits	1.042	↑	No	1.078	↑	No	1.014	↑	No
	Social Support	1.023	↑	No	.839	↓	No	1.013	↑	No
	Social Control	.921	↓	No	.868	↓	No	.903	↓	No
Black	Responsiveness to advice	1.219	↑	No	.967	↓	No	1.250	↑	No
	Prosocial Identity	.944	↓	No	1.459	↑	No	.869	↓	No
	High Expectations	1.128	↑	No	1.065	↑	No	1.239	↑	No
	Costs/Benefits	.884	↓	No	1.115	↑	No	.920	↓	No
	Social Support	1.013	↑	No	1.324	↑	No	1.061	↑	No
	Social Control	.940	↓	No	.558	↓	No	.891	↓	No
Hispanic	Responsiveness to advice	1.017	↑	No	--	--	--	.874	↓	No
	Prosocial Identity	2.503	↑	No	--	--	--	2.920	↑	No
	High Expectations	.791	↓	No	--	--	--	.917	↓	No
	Costs/Benefits	1.277	↑	No	--	--	--	1.625	↑	No
	Social Support	1.114	↑	No	--	--	--	1.171	↑	No
	Social Control	.733	↓	No	--	--	--	.425	↓	No

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry. Significant (Sig.) results are **bolded** and ↑ and ↓ describe the impact of each item on recidivism; ↑ indicates that likelihood of recidivism increases as scores on the item increase. – is used to identify results that could not be computed due to insufficient base rate.

The final set of analyses conducted in Study 3 sought to explore if any of the items in the DRAOR's Protective subscale functioned as true protective factors.[23] As defined by Jones et al., (2015), a protective factor is understood to interact with risk level, such that they are more salient for higher risk JIPs. Thus, the first step in identifying potentially protective factors was to test for interactions between items in the Protective subscale, and JIPs' risk level. Accordingly, interactions terms were created using item scores and risk. Analyses were conducted separately for technical violations and new offences, and results are presented in Table 45 and 46 below[24]. With respect to predicting technical violations, two protective items, high expectations and cost/benefit analysis demonstrated a significant interaction with risk, with $z = 8.49$, $p = .004$ and $z = 5.89$, $p = .015$ respectively for White women. None of the item by risk interactions significantly predicted technical violations for Black and Hispanic women. With respect to predicting new offences, two of item by risk interactions were significant; again, high expectations interacted with risk to predict recidivism for White women ($z = 5.76$, $p = .016$), and for Black women, the interaction between social support and risk reached significance ($z = 4.03$, $p = .045$). That said, only the high expectations by risk interaction for White women (for technical violations) remained significant following FDR adjustments to control for the familywise error rate following multiple comparisons. Thus, based on the presence of a significant interaction term, the high expectations item for White women was the only candidate protective factor, though it should be noted that the odds ratio greater than 1.0

[23] These analyses were undertaken with the understanding that identifying genuine protective factors was unlikely in view of the previous results involving promotive factors. Nevertheless, it was possible that such analyses would yield informative results regarding the interactions of item scores with risk.
[24] The any return outcome was omitted from these analyses as previous analyses revealed little difference between it and the technical violations outcome.

suggests that higher scores on the high expectations item appear to be associated with increased, rather than decreased, recidivism.[25]

[25] Problematically, despite the presence of a significant interaction with risk level, this item is an unlikely candidate for a protective factor as earlier analyses suggest that it may instead have an aggravating effect for White women.

Table 45

Interactions Between Protective Factors and Risk Level Predicting Technical Violations for White, Black, and Hispanic Women

	DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio
White	Responsiveness to advice x risk	.004	.010	.152	1	.697	1.004
	Prosocial Identity x risk	-.018	.010	3.320	1	.068	.982
	High Expectations x risk	**.027**	**.009**	**8.486**	**1**	**.004***	**1.028**
	Costs/Benefits x risk	**.024**	**.010**	**5.886**	**1**	**.015**	**1.024**
	Social Support x risk	.013	.009	2.182	1	.140	1.013
	Social Control x risk	.003	.011	.072	1	.789	1.003
Black	Responsiveness to advice x risk	.001	.021	.002	1	.963	1.001
	Prosocial Identity x risk	.004	.025	.030	1	.862	1.004
	High Expectations x risk	.005	.022	.057	1	.812	1.005
	Costs/Benefits x risk	-.020	.023	.711	1	.399	.980
	Social Support x risk	.019	.023	.712	1	.399	1.020
	Social Control x risk	.013	.026	.272	1	.602	1.013
Hispanic	Responsiveness to advice x risk	-.045	.050	.809	1	.368	.956
	Prosocial Identity x risk	.058	.050	1.327	1	.249	1.060
	High Expectations x risk	.018	.043	.171	1	.679	1.018
	Costs/Benefits x risk	.018	.044	.170	1	.681	1.018
	Social Support x risk	.020	.041	.236	1	.627	1.020
	Social Control x risk	-.013	.061	.044	1	.834	.987

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry, **bold** denotes significance at the $p < .05$ level, **bold*** denotes significance after FDR correction for multiple tests.

Table 46

Interactions Between Protective Factors and Risk Level Predicting New Offences for White, Black, and Hispanic Women

	DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio
White	Responsiveness to advice x risk	-.011	.014	.650	1	.420	.989
	Prosocial Identity x risk	-.017	.015	1.307	1	.253	.983
	High Expectations x risk	**.030**	**.013**	**5.763**	**1**	**.016**	**1.031**
	Costs/Benefits x risk	.018	.014	1.866	1	.172	1.019
	Social Support x risk	-.008	.013	.366	1	.545	.992
	Social Control x risk	-.005	.015	.089	1	.766	.995
Black	Responsiveness to advice x risk	-.009	.029	.108	1	.743	.991
	Prosocial Identity x risk	.029	.032	.812	1	.368	1.029
	High Expectations x risk	-.003	.029	.009	1	.926	.997
	Costs/Benefits x risk	.010	.031	.111	1	.740	1.010
	Social Support x risk	**.055**	**.028**	**4.034**	**1**	**.045**	**1.057**
	Social Control x risk	-.033	.033	1.035	1	.309	.967
Hispanic	Responsiveness to advice x risk	.106	.083	1.620	1	.203	1.112
	Prosocial Identity x risk	.039	.094	.172	1	.679	1.040
	High Expectations x risk	-.048	.075	.420	1	.517	.953
	Costs/Benefits x risk	-.117	.098	1.411	1	.235	.890
	Social Support x risk	.081	.071	1.283	1	.257	1.084
	Social Control x risk	-.015	.110	.020	1	.889	.985

Note. DRAOR = Dynamic Risk Assessment for Offender Re-Entry, **bold** denotes significance at the $p < .05$ level, **bold*** denotes significance after FDR correction for multiple tests.

In theory, the final step in identifying a protective factor involves examining the predictive ability of the item at each level of risk. Unfortunately, these analyses were hampered by limitations in the data (as well as conflicting earlier results) and conclusions here are largely speculative. As demonstrated in by the non-significant results and directions of the odds ratios shown in Table 47, the high expectations item does not function as a protective factor against technical violations for White women. However, there is some ambiguity in results as it was not possible to complete analyses for the high risk group; small sample size at this level of disaggregation produced highly unstable and uninterpretable regression coefficients. As protective factors are expected to be most salient for high risk JIPs, this could be seen as a serious limitation. Available results indicate that scores on the high expectations item did not significantly predict technical violations for those assessed as low-moderate, moderate, and moderate-high risk, and as noted earlier, the observed odds ratios are greater than 1.0 (rather than below 1.0), suggesting that high scores on this item are actually associated with increased likelihood of technical violations. In sum, these results are concerning as high expectations appear to have an aggravating rather than protective influence for moderate to high risk White women but should be viewed cautiously as a number of limitations affected these analyses.

Table 47

Regression Analyses Examining Predictive Ability of the High Expectations Item for Low-Moderate, Moderate, and Moderate-High Risk Women Predicting Violations

Risk Level	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio
Low-Moderate	.084	.162	.268	1	.605	1.087
Moderate	.115	.153	.570	1	.450	1.122
Moderate-High	.074	.140	.281	1	.596	1.077

Discussion

The overarching goal of Study 3 was to comprehensively evaluate the DRAOR's ability to predict various recidivism outcomes and to determine if these abilities varied according to JIPs' gender and race. As a risk assessment instrument, the DRAOR is intended to assist parole officers in assessing and monitoring the level of risk posed to the community by the JIPs they supervise (Serin, et al., 2016). To date, little research has evaluated how DRAOR subscale and item scores differentially predict various types of recidivism for men and women JIPs of different races. As prescribed by the responsivity principle of the RNR model (Andrews et al., 2016), an empirical understanding of how DRAOR subscale and item scores function across gender and race is required for parole officers to successfully tailor supervision strategies to individual JIPs.

The expressed intent of this study was to explore the extent to which intersectionality (i.e., the conceptualization that interconnected social categorizations such as race and gender create overlapping and interdependent systems of discrimination or disadvantage; Crenshaw, 1990) might be apparent in the correctional context of risk assessment for JI women, and Black and Hispanic JI women in particular. To explore

this, the DRAOR's predictive ability was assessed by means of discrimination (relative predictive accuracy; *AUCs*), calibration (absolute predictive accuracy; E/O index statistics), and predictive modeling (regression; odds ratios). Results were disaggregated by gender and race. This study also investigated whether individual DRAOR items exerted a promotive and/or protective effect for JI women.

Summary of Findings

Differences in means. Justice involved women had higher DRAOR Total scores (M = 7.2, SD = 7.2) than men (M = 7.0, SD = 6.8). Compared to men, women received higher scores on each of the subscales, though their scores were only significantly higher on the Acute and Protective subscales. This would suggest that the women in the sample were assessed as higher risk, but also as having more potentially mitigating factors present. These findings are largely consistent with those of Yesberg, Scanlan, and Polaschek (2014) who, in a similar study, compared DRAOR scores in a sample of 145 women and 145 case-matched men. However, in their sample, the women received lower Protective scores. Overall, the fact that women received higher scores on the Stable and Acute subscales, which capture risk, was not unexpected and corresponds to the finding that women also had higher rates of technical violations.

With respect to mean scores on individual items, findings were largely consistent with expectations based on previous research (eg., Benda, 2005; Brown, 2017; Cimino et al. 2015; Gobeil et al., 2016; Kopak et al., 2015). For example, women were rated as having greater need on items related to personal/emotional circumstances (e.g., problem solving, negative mood), relationships (e.g., peer associations, interpersonal relationships), and employment. Interestingly, women were not rated as higher risk on the

substance abuse item, though that does not necessarily mean that the item was not salient for women.

Differences were also evident across race. Black JIPs had the highest DRAOR Total scores of the three race groups, including significantly higher Stable scores when compared to White JIPs, though not Hispanic JIPs. These results correspond to the base rates observed in this sample (see Study 2) and were therefore expected. Black JIPs had the highest mean scores on several items, including attitudes toward authority, sense of entitlement, anger, and employment, and White JIPs scored highest on substance abuse, negative mood, and interpersonal relationships. However, White JIPs were also more likely to receive higher social support ratings when compared to Black and Hispanic JIPs. Hispanic JIPs had the lowest DRAOR scores overall and were not found to have significantly higher mean scores on any of the DRAOR items when compared to their Black and White counterparts.

Predictive accuracy of DRAOR Total and subscale scores.

Discrimination. Relative predictive accuracy was examined separately for each of the three recidivism outcomes and discrimination was compared across gender and race. Somewhat contrary to expectations, DRAOR subscale and Total scores did not demonstrate adequate (or comparable) predictive accuracy for each of the subgroups examined. For technical violations and the any return outcome, DRAOR scores demonstrated adequate discrimination for White men and women (*AUC*s between .56 and .61), but there was considerably more variability in the relative predictive accuracy observed for Black and Hispanic men and women. Discrimination was poorest for Black JIPs, and men in particular. For Black men, DRAOR Total and subscale scores predicted

technical violations and returns at a level analogous to chance (*AUC*s between .49 and .54), and discrimination was only marginally improved for Black women (*AUC*s between .52 and .55). The discrimination offered by DRAOR scores was most variable for Hispanic JIPs, with *AUC*s ranging .54 (not significantly different from chance) to .73 (excellent discrimination). Notably, while discrimination was consistently higher for Hispanic men than Hispanic women, the same pattern in terms of strength of prediction was observed across DRAOR Total and subscale scores. Stable scores consistently provided the best discrimination (*AUC* = .73 and .73 for technical violations and any return for Hispanic men and *AUC* = .63 and .65 for technical violations and any return for Hispanic women), followed by DRAOR Total scores discrimination (*AUC* = .67 and .69 for technical violations and any return for Hispanic men and *AUC* = .61 for both technical violations and any return for Hispanic women). Discrimination was significantly lower for Protective and Acute subscale scores, with *AUC*s ranging from .56 to .63 for men, and .53 to .55 for women, though it is worth noting that these *AUC*s were considerably better than the *AUC*s obtained for Black JIPs and comparable to those obtained for White JIPs.

Overall, DRAOR Total and subscale scores demonstrated poor relative predictive accuracy with respect to the new offence outcome. Across all gender and race groups, *AUC*s ranged from .40 to .60, with an average *AUC* of .536. Interestingly, for the new offence outcome, discrimination was highest for Black women. DRAOR Total, Stable, and Acute scores modestly discriminated recidivists from non-recidivists, with *AUC*s of .58, .59, and .58, respectively. Discrimination was similar though slightly poorer for White men (*AUC*s of .56, .57, and .55 for DRAOR Total, Stable, and Acute scores), and

results were weaker still for White women and Black men. DRAOR scores did not provide reliable discrimination for Hispanic JIPs.

Considered collectively, results indicated that relative predictive accuracy differed more as a function of race than gender. All *AUC* difference tests evaluating whether or not discrimination differed across gender yielded non-significant results whereas several of the comparisons made between racial groups pointed to meaningful differences[26]. For instance, when predicting technical violations, both DRAOR Total and Acute scores had higher predictive accuracy for White men and women compared to Black men and women. Moreover, Stable scores were better able to differentiate between individuals who did and did not incur another technical violation for Hispanic men than they were for White and Black men. Though Stable scores also demonstrated good predictive ability for Hispanic women, discrimination was not meaningfully improved for Hispanic women relative to White and Black women. Results regarding differences in discriminative ability were largely consistent when predicting any returns for JI men, though, for this recidivism outcome, DRAOR Total and Acute scores did not provide improved discrimination for White women when compared to Black women. No significant differences in discriminative ability were found between race and gender groups for the prediction of new offences.

Calibration. The E/O index statistic was used to assess calibration, and the expected number of recidivists in the current study (the 'E' in the E/O index) was

[26] It is important to note that a significant difference test does not imply that the scores do not differentiate between recidivists and non-recidivists meaningfully for both groups. Rather, a significance difference test suggests that the predictive accuracy may be significantly higher for one group (e.g., White individuals) compared to another (Black individuals)

calculated based on data describing 343 White men serving community supervision orders in Iowa collected during an earlier pilot study (see Chadwick, 2014). Initially, this study attempted to assess calibration for individual DRAOR items. However, inadequacies in sample size for the smaller race groups (i.e., Black and Hispanic JIPs), coupled with low base rates for the new offence outcome resulted in a revised approach. Instead, calibration was assessed at three levels for each DRAOR subscale (low, mean, and high) and at four levels (low-moderate, moderate, moderate-high, and high) for DRAOR Total scores.

Analyses at this level revealed that DRAOR scores were not consistently well calibrated to predict recidivism for many of the gender by race subgroups examined in this study. For example, despite the fact that the number of expected recidivists was derived from a sample of presumably similar White male JIPs also supervised in Iowa several years earlier, DRAOR Total and subscale scores did not provide consistent levels of absolute predictive accuracy across the technical violation, new offence, and any return outcomes for the White men in this sample. High Stable scores demonstrated good calibration when predicting technical violations (E/O = 1.05), but poor calibration when predicting new offences (E/O = 2.42). Collectively, calibration results for White men suggested that technical violations and returns were under-predicted for those assessed as lower risk (i.e., those with low Stable and Acute scores, and high Protective scores), and over-predicted for men assessed as high risk (i.e., those with high Stable and Acute scores, and low Protective scores). This pattern was also observed for Black men. While more of the E/O index statistics were non-significant (and therefore theoretically indicative of good calibration) when compared to White men, the extent of under- and

over-prediction of recidivism was likewise more pronounced. Stable scores appeared to demonstrate the best absolute predictive accuracy for Black men, though in many cases the ratio of expected to observed recidivists was quite far from 1.0 and many of the confidence intervals associated with the E/O indexes were imprecise. For Hispanic men, results suggested that calibration was reasonable overall, with over- and under-predictions tending to be by relatively small proportions. However, the small sample size resulted in highly unstable confidence intervals and results should be viewed as exploratory.

Calibration results for women were similar to those described above for men in that a similar progression from under-prediction at lower levels of risk to over-prediction at highest levels of risk was observed. Interestingly, absolute predictive accuracy appeared to be slightly better for White women as compared to White men, with the E/O indexes predicting technical violations and any return from Stable, Acute, and Protective scores being more closely clustered around 1.0. For Black women, calibration was significantly better for those rated as higher risk; E/O index statistics for Black women with high Stable and Acute scores and low Protective scores approximated 1.0 (range = .93 to .98) whereas for low risk Black women (low Stable and Acute scores, high Protective scores), recidivism was significantly under-predicted (range = .44 to .58). Calibration results for Hispanic women closely mirrored those seen for Hispanic men (i.e., theoretically non-significant E/O index statistics combined with very broad and imprecise confidence intervals) and were also consistent the trend observed for White and Black women where calibration improves as risk level increases.

Considered collectively, results of calibration analyses for low, mean, and high DRAOR subscale scores suggest calibration varies considerably as a function of risk level. While differences in the absolute predictive accuracy of DRAOR scores were present at the level of gender and race in the current sample, level of risk also appears to have a significant impact. In the current sample, higher Stable and Acute scores and lower Protective scores demonstrated better calibration, and recidivism was routinely under-estimated in JIPs with lower DRAOR scores. Importantly, the smaller sample sizes for Black and Hispanic JIPs negatively impacted the precision of results. Thus, conclusions about how calibration may vary across race would be premature.

To increase the stability and interpretability of results, a series of overarching calibration analyses assessing the absolute predictive accuracy of DRAOR scores predicting each of the three recidivism outcomes were conducted for each gender by race subgroup. Results indicated that DRAOR Total scores were well calibrated to predict technical violations for Black and Hispanic men, and Hispanic women. Interestingly, DRAOR Total scores were not well calibrated for White JIPs for technical violations. For the prediction of new offences, evidence of good calibration was only observed for Black men, though the computed E/O index (1.73) was associated with wide confidence intervals and indicates that new offences were over-estimated considerably, if not significantly. New offences were over-estimated to an even greater extent for all other groups of JIPs. Across the six gender by race groups examined in this study, calibration was best for the any return outcome. Good calibration was observed for DRAOR Total scores for White, Black, and Hispanic men, and for Hispanic women. Returns were

slightly underestimated for White women (E/O = .92) and though the confidence intervals did not contain 1.0, they were quite close (lower CI = 0.87, upper CI = 0.97).

Importantly, absolute predictive accuracy was quite good for the technical violations and any return outcomes when assessed at the level of overall DRAOR Total score (i.e., not subdivided by subscale). When examined within individual subscales and at the item level, calibration was inconsistent at best. This disparity in predictive accuracy suggests that results within subscales, and certainly at the item-level, were significantly, and negatively, impacted by sample size and low base rates (for the new offence outcome). These results also demonstrate real differences in the base rates of the technical violations, new offences, and returns across the reference sample and the current sample. Persons in this sample were slightly more likely to incur technical violations and returns, but significantly less likely to commit a new offence.

Contextualizing relative and absolute predictive accuracy findings. Overall, results of the discrimination and calibration analyses conducted to assess predictive accuracy do not support the first hypothesis made in Study 3. Contrary to expectations, differences in predictive accuracy were observed across the six race by gender study groups, and predictive accuracy, though excellent in some cases (e.g., the relative predictive accuracy of Stable scores for Hispanic men), was very poor in others. These results, and those describing the relative predictive accuracy of DRAOR Total and subscale scores in particular, are somewhat difficult to reconcile with prior research. Previous research on the DRAOR (see Table 1) has generally found that DRAOR demonstrates adequate discriminative validity *AUCs* ≈ .60) across various subpopulations, including men, women, youths, violent JIPs, sex offenders, and

Indigenous (Maori) JIPs. Though many of the *AUC*s calculated in this study met or exceeded that threshold, many (e.g., DRAOR Total scores predicting technical violations for Black men [*AUC* = .50]) also fell far below it. Findings from the current study are also at odds with those of a recent study by Serin et al., (2018), which suggested that both discrimination and calibration were adequate for White, Black, and Hispanic JI men and women in Iowa. However, Serin and colleagues (2018) were not able to disaggregate by gender and race simultaneously, which may account for some of the difference seen in the present study.

Logistic regression analyses. Logistic regression was used to explore the predictive ability of individual DRAOR items for White, Black, and Hispanic men and women. A main advantage of regression analysis is that it produces an odds ratio (*OR*) statistic, which provides an index of the strength of the association between scores on a given item and the outcome of interest. As the overarching goal of this dissertation was to contribute to the continued integration of the fields of risk assessment and offender management while attending to intersectionality, evaluating how changes in DRAOR scores impact the likelihood of recidivism for men and women of different races was a central goal.

Prior to exploring prediction at the item level, a series of omnibus regression models predicting recidivism from the set of Stable, Acute and Protective scores were run. Already at this level, it became apparent that the second primary hypothesis made in Study 3, that prediction would vary as a function of gender but not race, would not be supported. The majority of the variability in results was observed across racial groups,

though small differences were also seen across gender. DRAOR scores reliably predicted recidivism for White JIPs but were poor predictors for Black and Hispanic JIPs.

The set of subscale scores entered as covariates significantly predicted technical violations, new offences and any return for White men and technical violations and any return for White women. The model approached significance for Hispanic men when predicting technical violations and any return but did not maintain significance following the required adjustments to significance levels when using multiple tests. The set of Stable, Acute, and Protective scores did not reliably predict recidivism for Black JIPs. That said, even in the significant models, the set of DRAOR covariates did not account for much of the variance in outcome, with all observed McFadden's ρ^2 values falling far below the 0.2 benchmark commonly cited for a small effect (Hensher & Johnson, 1981).

The impact of subscale scores was also assessed separately for each gender by race group. Stable, Acute, and Protective scores significantly predicted technical violations and any return for White men and women, and odds ratios indicated that increases in scores on each subscale were associated with changes in the likelihood of recidivism in the intended direction. For example, for White women, a one point increase in Stable, Acute, and Protective subscale scores were associated with odds ratios of 1.13, 1.13, and .91, respectively. Stable, Acute, and Protective scores did not reliably predict recidivism for Black men or women, regardless of the recidivism outcome examined and results suggested that overall, DRAOR scores were least useful for Black JIPs. Though the predictive ability of DRAOR subscale scores for Hispanic JIPs was also limited, Stable scores did reliably predict technical violations and returns for Hispanic men, with odds ratios of 1.42 and 1.40 respectively. This finding corresponds with the results of

earlier discrimination analyses, where Stable scores demonstrated excellent relative predictive accuracy (i.e., *AUC*s > .70) when predicting violations and returns for Hispanic men. Notably, subscale scores did not significantly predict new offences for any of the subgroups with the exception of Stable scores for White men. However, closer examination of the odds ratio associated with this model (*OR* = 1.01) reveals that despite achieving significance, these results provide little information of practical significance to case managers.

Item-level prediction. Separate models predicting technical violations, new offences, and any return from individual DRAOR item scores were analyzed for each of the six race by gender groups. Of the 19 DRAOR items, only three (interpersonal relationships, living situation, and high expectations) significantly predicted technical violations, and only two (substance abuse and interpersonal relationships) reliably predicted returns. Moreover, these items were only significant predictors for White and Black men; no individual items reliably predicted technical violations, new offences, or returns for women, regardless of their race. The finding that interpersonal relationships, living situation, and high expectations were reliable predictors for Black men was somewhat unexpected given that none of the subscale scores emerged as significant predictors, but it does illustrate the need for item-level analyses. Though it is slightly disappointing that no DRAOR items reliably predicted outcome for JI women given the goals of this study, these findings are still informative.

For example, while they did not reliably predict recidivism following adjustments made to control familywise error rate, several items did appear to have some salience for women, including substance abuse, employment, and prosocial identity for White

women, and attitudes toward authority for Black women. These findings are somewhat consistent with hypotheses, as previous research (e.g., Cimino et al. 2015; Gobeil et al., 2016; Kopak et al., 2015) has found associations between substance use and personal/emotional factors and recidivism in women. Similarly, while substance abuse appears to bear some relationship (if not always a significant one) to technical violations and returns for White men and women, this relationship does not appear for Black women, or Hispanic men or women. This has interesting implications for supervision. It is also interesting that interpersonal relationships and living situation are more strongly related to violations and returns for men; interpersonal relationships in particular were expected to be more salient for women. With respect to intriguing findings related to the intersection of gender and race, attitudes toward authority demonstrated some relationship to offending for Black women and Hispanic men, but not for other groups. Notably, no DRAOR items were reliably associated with recidivism (of any kind) for Hispanic women, and White women were the only group for whom DRAOR scores appear to bear any relationship with new offences.

Considered collectively, these findings suggest that few DRAOR items reliably predict recidivism, even for White men. Moreover, though these findings do not clearly identify protective factors for women or individuals belonging to historically marginalized groups, they do suggest that prediction may be even more limited for these groups. Clearly, the hypothesis that the predictive ability of DRAOR items would be consistent across race was not supported.

Promotive and protective factors. Whether the six items comprising the Protective subscale of the DRAOR functioned in a promotive or protective capacity was

examined for White, Black, and Hispanic JI women. Results were highly inconsistent with expectations. Only one item, prosocial identity, was found to be significantly associated with reductions in recidivism, and moreover, this effect was only present for White women in relation to technical violations and any returns. Concerningly, results suggested that higher scores of some of the supposed protective factors appeared to be related (though not significantly) to increases in recidivism, rather than decreases. Additionally, which items produced an increase in likelihood of recidivism (i.e., *OR*s > 1.0) varied according to race and in relation to the recidivism outcome being considered. For instance, the high expectations item was associated with a (non-significant) increase in likelihood of technical violations and new offences and a significant increase in the likelihood of any return for White women, with a non-significant increase in likelihood of recidivism for all three outcomes for Black women, and with a non-significant decrease in likelihood of technical violations and any returns for Hispanic women. Clearly, there is a need for further research examining how JI women are being rated on the DRAOR's protective items.

The final set of analyses in Study 3 sought to explore whether any of the items in the Protective subscale functioned as genuine protective factors according to the definition put forth by Jones and colleagues (2015). Notwithstanding the fact that only prosocial identify was found to have a promotive effect, it was still of interest to test whether the salience of some of these items might vary as a function of risk though at this point, meaningful results were not anticipated. Interactions were tested between individual DRAOR item scores and women's level of risk, and unsurprisingly, only one interaction term, high expectations x risk predicting technical violations for White

women achieved significance. Problematically, earlier analyses indicated that the high expectations item could more accurately be considered an aggravating factor for White women. No evidence was found to suggest that this item was more or less salient for White women with low-moderate, moderate, or moderate-high DRAOR Total scores, but results did definitely demonstrate that this item did not have a protective effect for White women in the current study.

Possible Explanations of Findings

The results of Study 3 differed significantly from findings of previous examinations of the DRAOR and related research on dynamic risk and protective factors, and therefore also from expectations. Though not all of the results in the current study can be easily explained, there are nevertheless several factors that are expected to have contributed to results. First, it must be noted that it is possible that the Black and Hispanic JIPs included in this sample simply have risk and/or need factors not fully captured by the DRAOR. Though this may be a contributing factor, is unlikely that this is the primary factor driving results. The DRAOR was developed for general correctional samples such as this one, and the results of previous studies using the DRAOR, some of which were also conducted in Iowa on earlier samples, suggest that these findings are more exception than rule. Notably, pilot studies in Iowa (e.g., Chadwick, 2014; Serin & Prell, 2012) found that DRAOR scores discriminated recidivists from non-recidivists with moderate accuracy (*AUC*s between .55 and .74). Acceptance of this explanation would imply that the correctional population in Iowa changed substantially within the span of a few years.

Second, this study attempted to disaggregate analyses to a greater extent than prior DRAOR research. While previous studies have disaggregated analyses by gender

(e.g., Yesberg et al., 2015; Scanlon, Fortune, Polaschek & Yesberg, 2015) or race (e.g., Hanby, 2013) individually, to my knowledge, this is the first study large-scale study that attempted to simultaneously disaggregated by gender and race while also attempting to explore prediction within levels of risk and at the item level. Accordingly, it is reasonable to expect that findings would differ on account of the variation in level of analysis. Relatedly, disaggregating to a greater extent reduces sample size. Prediction analyses such as discrimination, calibration, and logistic regression are heavily reliant on sample size (Tabachnick & Fidell, 2013), and the fact that the Black and Hispanic samples were not only relatively small, but also considerably smaller and more variable than the sample of White JIPs undoubtedly had an impact on results.

Third, findings from the current study can be at least partially explained by the results of Study 1. The DRAOR did not demonstrate measurement invariance for White, Black, and Hispanic women. While Study 1 did not examine the stability and consistency of the DRAOR's factor structure across men of different races, it is likely that similar inconsistencies would have been detected. A lack of measurement invariance precludes comparisons between groups as it cannot be assumed that DRAOR scores mean the same thing for JIPs of different races. In other words, if both a White woman and a Hispanic woman receive a score of 7 on the Acute subscale, these scores cannot be interpreted as having the same impact on likelihood of recidivism. In view of this it is hardly surprising that significant difference in predictive accuracy were detected across race.

Fourth, this study relied on baseline DRAOR assessment scores to predict later recidivism. As such, it is possible that supervision staff may have met with their clients and modified supervision strategies one or more times between the initial assessment and

any outcome. The current study was not able to control for the potential influence of appropriate supervision responses (i.e., reductions in recidivism) which could have contributed to limited predictive accuracy. As discussed earlier, parole officers are often pressed for time (e.g., Adelman, 2020; Matz et al., 2018) and may not have enough time during supervision meetings, and initial supervision meetings in particular, to gather the information required to accurately score each DRAOR item. Many items (e.g., problem solving skills, responsiveness to advice) require substantial background information and understanding of an individual's circumstances which may be difficult to collect in a short time, especially if the quality of the working relationship between the parole officer and client is lacking. Baseline assessment scores were used in the current study; thus, parole officers may have been meeting the individual for the first time. As noted by Lloyd et al., (2020), most proximal assessment scores, which yielded the highest predictive accuracy, likely incorporated implicitly averaged information about the individual's risk and need gathered over the course of prior meetings.

Finally, there is likely variation in how parole officers score DRAOR items and arrive at Total scores. Though scoring manuals are available and standardized training and credentialing is a requirement, it may still be the case that there is variation among supervision officers in the use of the DRAOR. Other external factors (e.g., the quality and recency of their training) as well as internal factors (e.g., their level of conscientiousness and commitment to their job; Hanson et al., 2007) could also impact scoring fidelity, which in turn would affect predictive validity. Degree of "buy in" among supervision officers is also unknown. The level and amount of staff buy in can have a profound impact on the efficacy of an assessment measure (Latessa & Lovins, 2010; Schlagger,

2009). Staff members who are not convinced of the utility of a measure may be less inclined to complete it diligently. Importantly, they may also be less likely to use the scores to inform decision-making (Miller and Maloney, 2013). With respect to the disconcerting findings regarding the lack of promotive and protective effect demonstrated by items in the Protective subscale, many scholars (e.g., Schlager, 2018; Fortune & Ward, 2017) have claimed that strengths and protective factors are not accorded an appropriate amount of importance in the criminal justice system, including by front-line staff. There has been an ongoing debate in scholarly circles regarding appropriate definitions, measurement strategies, and practical applications of protective factors (see, for example, Serin et al., 2016; Ward, 2017; Ward & McDonald, 2016).

Implications for Practice

The results of Study 3 have a range of implications for correctional practice. However, given concerns about scoring fidelity and the fact that this study relied solely on baseline assessment scores, it should be noted that there may be generalizability issues for correctional agencies outside of the Iowa.

More broadly, findings suggest that particular care is required when using the DRAOR with JI women and historically marginalized groups. Women in the sample consistently received higher DRAOR scores (both overall and on individual subscales and items) than their male counterparts, which would suggest that JI women present a greater risk than JI male. However, examination of recidivism rates clearly reveal that women incurred fewer technical violations following release and were considerably less likely to commit a new offence. As such, parole officers should be aware that while changes in DRAOR scores may be meaningful for monitoring changes in individual JIPs,

that comparisons should not be made across JIPs on the basis of DRAOR scores, especially if race and gender differences are present. Markedly, this conclusion is underscored by the results of Study 1, which demonstrated that the DRAOR did not achieve measurement invariance for JI women of different races.

Relatedly, results of this study suggest that the predictive accuracy afforded by DRAOR scores varied more as a function of race than gender. Prediction was consistently strongest for White JIPs and weakest for Black JIPs but, was not consistently better for men compared to women of the same race. In fact, in some cases (e.g., the prediction of new offences from DRAOR Total scores for Black JIPs) prediction appeared to be better for women. Collectively, results of prediction analyses support the use of the DRAOR with White JIPs to predict technical violations and any return but suggest that DRAOR scores do not reliably predict these outcomes for Black and Hispanic JIPs, with the notable exception of scores on the Stable subscale for Hispanic men. Alternative and/or adapted approaches to risk assessment are therefore recommended for Black and Hispanic JIPs until causes underlying the poor performance of the DRAOR with these groups are identified and addressed. The extent to which these findings can be attributed to intersectionality remains unclear. Of the recidivism outcomes examined, DRAOR scores predicted new offences least well for all JIPs. However, as similar results had been found in previous examinations of the DRAOR (e.g., Chadwick, 2014; Serin et al., 2018) this was not surprising.

Finally, differences in predictive ability were also detected as a function of risk such that predictive ability was better for higher risk individuals. Calibration analyses demonstrated that likelihood of recidivism was consistently under-estimated for lower

risk JIPs (i.e., those with lower Stable and Acute scores and higher Protective scores) and it therefore suggested that case managers continue to bear in mind the responsivity principle of RNR (Andrews et al., 2016) when making decisions about how to best allocate their time and resources.

Limitations and Future Directions

Findings from the present study should be considered in light of several important limitations. First, despite seemingly large sample size, the number of Black and Hispanic JIPs in the sample was insufficient given the intended level of disaggregation. As discussed in earlier sections, discrimination, calibration and regression analyses are susceptible to small sample size and this issue was exacerbated by item- and subscale-level analyses. This was even more problematic when examining JIPs at extremes of the risk continuum (i.e., those with very high or low DRAOR scores) and for the new offence outcome, which was affected by low base rates (see Tabachnick & Fidell, 2013 for a description of how low base rates affect prediction analyses).

Second, the calibration analyses conducted in this study were limited by the lack of a true normative sample. Ideally, the expected number of recidivists forming the numerator of the E/O index would be calculated based on a set of empirically derived and validated norms (i.e., recidivism rates; see Helmus, Hanson & Thornton, 2009 for a discussion of norm development with the Static-99). As no such norms currently exist for the DRAOR, this was not possible. Though the use of the Iowa pilot data (Chadwick, 2014) as a reference sample for calibration analyses in this study is justifiable, the lack of validated norms limits the replicability and generalizability of results. The results of this

study caution against further calibration analyses until validated norms have been developed for the DRAOR.

Third, as discussed in Study 2, how the outcome variables (i.e., technical violations, new offenses, and any return) were defined represent another important limitation. There was a considerable degree of variability in each of these categories (e.g., both petty theft and murder would be considered a new offence) and this variability likely affected base rates, which, by extension, contributed to the inconsistent predictive accuracy seen across groups, especially for Black and Hispanic JIPs for whom small sample size was also an issue. More equitable grouping of outcome variables is advised.

Fourth, no information about correctional staff completing DRAOR assessments was available for this study. Though all staff in Iowa receive training, there are other potentially influential staff variables (e.g., level of experience, training, belief in the utility of the DRAOR) that could have impacted rating fidelity and how assessments were completed for individuals on their caseload (see Schaefer & Williamson, 2018; Viljoen et al., 2019). Relatedly, information pertaining to operational variables, such as training quality and compliance monitoring across IDOC jurisdictions was likewise unavailable but seemingly important (Chadwick, 2014). The concerning findings regarding the items in the Protective subscale (i.e., that many appeared to function as aggravating factors) points to a real need to examine scoring practices.

Future research on the DRAOR should seek to address the limitations described above. To rigorously assess the predictive ability of risk assessment tools like the DRAOR while disaggregating by several characteristics (e.g., race, gender, risk level) samples sizes of 1,000 or more (for each subgroup of interest) are recommended,

especially if item-level analyses are of interest. Likewise, careful consideration of how outcome variables will be defined is advised. The goal should be to strike a balance between having outcome variables that are broad enough to ensure that low base rates are not problematic, but not so broad that there is considerable variability in the severity nature of the outcome, as was the case in the current study.

The current study also highlights the need for empirical development and validation of recidivism norms for the DRAOR. Such norms would not only facilitate future calibration analyses but would also greatly improve researchers' ability to communicate meaningfully about JIPs assessed using the DRAOR and contribute to the standardization of assessment and supervision practices more broadly. Also daunting, results of the current study confirm the need to revisit the composition of the DRAOR. Study 1 raised questions regarding the appropriateness of the DRAOR's organization according to the current three factor structure, and results of this study raise further questions about the appropriateness of individual items. Additional research is warranted as it is unclear whether the observed issues stem from the tool itself or idiosyncrasies of the current sample.

Finally, from an implementation perspective, future DRAOR research (both in Iowa and more generally) should aim to explore the rating fidelity of case manager assessments. Specifically, examinations of (a) the quality and availability of DRAOR scoring training, (b) inter-rater reliability, (c) staff buy-in and compliance monitoring, (d) the impact of case manager – client relationships, and (e) how DRAOR scores are used to inform intervention strategies is recommended.

Conclusion

The current study suggests that DRAOR scores reliably predict technical violations and any returns for White men and women. With the exception of Stable subscale scores for Hispanic men, DRAOR scores did not reliably predict recidivism for Black and Hispanic men and women.

Prediction of new offences was poor for all gender by race groups, including White men. Though small differences in predictive accuracy were observed between men and JI women belonging to the same racial group, predictively accuracy was more strongly influenced by race (i.e., worse for JI individuals belonging to historically marginalized groups). Also, JIPs' risk level was found to have an impact on predictive accuracy, such that DRAOR scores predicted outcome more accurately for higher risk individuals. Recidivism was consistently under-estimated for JIPs assessed as lower risk on the DRAOR.

Individual DRAOR items scores contributed little to the prediction of recidivism, even for White men and women. That said, as only baseline DRAOR scores were examined, this study is not in a position to speculate about how changes in item scores might be used to inform case management practices. Importantly, the current study raises a number of questions with respect to items comprising the Protective subscale. Contrary to expectations, the DRAOR's presumed protective factors did not function as protective factors for White, Black, and Hispanic women in the sample, and many did not even exert a significant promotive effect. Instead, results indicated that some items actually had an aggravating effect.

Clearly, additional research exploring how prediction is impacted by not only gender but also race is merited. While the findings from the current study did not provide clear support for the gender- or race-salience of individual DRAOR items, many of the items comprising the DRAOR (e.g., substance abuse, interpersonal relationships and peers, living situation and employment) have been found to fulfill this function in other research (e.g., Brown, 2017; Cimino et al. 2015; Gobeil et al., 2016; Kopak et al., 2015; Olver et al., 2014). Similarly, research such as Jones and colleagues (2015) and Lodewijks et al., (2010) speaks to the promotive and protective effects of factors such as social support and high expectations for success. It is important that this research continue. Without discounting the results of the current study, it is possible that these findings represent an exception rather than the rule. To more conclusively establish if present findings are due to implementation issues such as scoring fidelity rather than actual characteristics of the Iowa correctional population replication research would be required.

Chapter 8: General Discussion

Accurate risk assessment and all that it entails has preoccupied correctional agencies and researchers for decades. Offender assessment pervades correctional systems internationally and plays a critical role at all junctures of the correctional process. Historically, the development, validation, and application of risk assessment tools has focused on White males. However, the need to expand these practices to better address the needs of other rapidly growing segments of the correctional population, namely women and historically marginalized groups is now being recognized. Equally, recent

decades have seen increased acknowledgment of the need to integrate the heretofore largely parallel research fields of dynamic risk assessment, case management, and offender desistance. Accordingly, a central goal of this doctoral research was to address risk assessment, risk management, and offender desistance attending specifically to the influence of gender and race. To do this, this dissertation undertook an examination of the Dynamic Risk Assessment for Offender Re-Entry (DRAOR; Serin, 2007), a case management tool comprised of dynamic risk and protective factors, in a large, racially diverse sample of JI women (N = 3,091).

Overview of Study Goals

Given the relative lack of research on dynamic risk and protective factors for women, and non-White women in particular, the first goal of this dissertation was to examine the factor structure and psychometric properties of the DRAOR with this unique population. To comprehensively address this question, several phases of analysis were required in Study 1. The first phase involved exploring the psychometric properties of the DRAOR and the extent to which the DRAOR's original three-factor structure provided a good fit to the data. Next, possible alternative models were tested. Once the best fitting model was identified, it was then subjected to invariance testing to determine if DRAOR scores could accurately be compared across race groups and over time. Statistical requirements for measurement invariance were not met. Thus, additional analyses were required to examine how individual DRAOR items contributed to the function of the assessment tool and whether refinement was possible.

Another area that warranted exploration was the relationship between JI women's race and recidivism rates. Previous research has raised concerns about the disproportional

representation of historically marginalized groups in correctional populations, and it was of interest to determine the extent to which recidivism rates differed for White, Black, and Hispanic women in the current sample. As such, the second goal of this dissertation (Study 2) was to examine JI women's failure rates in relation to three routinely documented correctional outcomes (i.e., technical violations, new offences, and any return). Whether survival time varied as a function of risk level for White, Black, and Hispanic women was also examined.

The final goal of this doctoral research was to comprehensively explore the predictive ability of DRAOR scores at the subscale level and at the level of individual dynamic risk and protective factors while disaggregating by race and gender. Though the predictive abilities of the DRAOR have been explored in previous research, to date, no research has specifically examined prediction for Black and Hispanic men and women or has sought to examine predictive ability of the DRAOR at the item level. Moreover, evaluations of predictive accuracy have typically relied on discrimination statistics (i.e., *AUC*s); in Study 3, the DRAOR's predictive ability was more comprehensively examined by means of discrimination, calibration, and regression. Finally, given the debate surrounding the concepts of gender-neutrality, gender-salience, and gender-specificity in contemporary correctional literature, whether any of the items contained in the DRAOR's Protective factor subscale functioned as promotive or protective factors for JI women merited investigation. Accordingly, how these items were related to recidivism for JI women was also examined.

Summary of Key Findings

Study 1. Study 1 evaluated the psychometric properties and factor structure of the DRAOR in a sample of White, Black, and Hispanic JI women (*N* = 2,591) on community supervision in Iowa. Previous research exploring the DRAOR's three-factor structure with diverse populations in Iowa (Chadwick, 2014) and New Zealand (Hanby, 2013; Lloyd, 2015) have yielded somewhat inconsistent results, and the current study was unable to provide clarification. Initial exploratory factor analyses (EFA) did not support the DRAOR's original structure and instead suggested that two alternative models, another three-factor structure with two different risk subscales and a four-factor structure with three risk subscales, would provide a better fit to the data. Notably, the Protective subscale emerged in its hypothesized (original) structure in both alternative models.

Importantly, for this dissertation it was less important that the DRAOR demonstrate a specific structure than a consistent structure. Informed by the results of the EFA, a series of confirmatory factor analysis (CFA) and exploratory structural equation modeling (ESEM) models were fitted. Consideration of relevant model fit indices and theory indicated the alternative three-factor CFA model provided the best fit for the data while also prioritizing conceptual parsimony. Thus, this model was used to examine measurement invariance across racial groups and over time. Measurement invariance, or factorial invariance across measurement occasions, refers to empirical evidence that scale items assess the same construct(s) across different measurement occasions. Multigroup confirmatory factor analyses (MGCFA) revealed that the DRAOR did not demonstrate the level of measurement invariance required to make meaningful comparisons across

race or over time (i.e., between first and last assessment). These findings have serious implications for how the DRAOR can be used in correctional settings.

Given the importance of measurement invariance for making group comparisons, supplementary analyses were undertaken in attempt to identify possible causes. Category response curves (CRCs) and item information curves (IICs) were generated to explore how individual items contributed to scale functioning. The goal of these analyses was to identify problematic items and explore whether it might be possible to remove said items. Unfortunately, only five of the 19 DRAOR items contributed significant information to the overall explanatory power of the scale, and the same five items (attitudes toward authority, sense of entitlement, attachment with others, anger/hostility, and access to victims) were the only items that meaningfully discriminated between scores of 0, 1, and 2. Considered collectively, results of Study 1 diverged quite strongly from those of prior DRAOR research, and called into question whether the instrument itself, characteristics of the current sample, scoring fidelity, or some combination of these factors led to the observed results.

Study 2. Study 2 explored difference in rates of recidivism and survival time for White, Black, and Hispanic women. Results indicated that recidivism rates did vary as a function of race, with Black women incurring more technical violation while on supervision than White and Hispanic women. Irregularities in base rates according to risk level were also detected. For example, for Black women, the recidivism rate did not increase monotonically in accordance with level of static risk as would be expected. Instead, Black women assessed as low risk incurred technical violations at rate comparable to those in the high risk group. Also concerning, White women in the lowest

risk group had a higher reoffence rate than their purportedly higher risk counterparts. That said, the sample size for the high risk group was very small (n = 34) and this may also have had an impact on findings.

Significant differences in survival time were also overserved across race. Rates of technical violations and any return were highest for Black women, though White women were found to have the shortest survival time when examining the new offences outcome. Analyses exploring the association between risk level and recidivism rates revealed that women's level of risk did have an impact on survival time, but again, results were inconsistent with expectations. Odds of failure did not increase proportionally with risk, and in several cases, the highest risk groups were found to have the lowest odds of failure. Overall, results of Study 2 demonstrate that there is a need to more carefully examine the routine supervision practices employed with JI women, and with criteria used to assign risk in particular.

Study 3. Study 3 aimed to explore the predictive abilities of the DRAOR while attending to intersectionality. The predictive ability of DRAOR scores was assessed by means of *AUC* statistics, E/O index statistics, and regression using matched samples of n = 2,763 men and women. Collectively, results of this study suggest that the predictive ability of DRAOR scores was more strongly impacted by individuals' race than gender. Significant differences in discrimination and calibration were evident across racial groups, and while significant differences across gender were anticipated, few were observed. Unsurprisingly, predictive accuracy was highest for White JIPs, and consistently poorest for Black JIPs.

Item-level analyses revealed that few DRAOR items significantly predicted recidivism, regardless of individuals' gender or race or the type of recidivism examined. Of the 19 DRAOR items, only three (interpersonal relationships, living situation, and high expectations) significantly predicted technical violations, and only two (substance abuse and interpersonal relationships) reliably predicted returns. Moreover, these items were only significant predictors for White and Black men – no individual items reliably predicted technical violations, new offences, or returns for women, regardless of race. Findings regarding the possible promotive and protective influence of the six items comprising the Protective subscale of the DRAOR were similarly equivocal. Only one item, prosocial identity, was found to be significantly associated with reductions in recidivism, and moreover, this effect was only present for White women in relation to technical violations and returns. Concerningly, results suggested that higher scores on some of the supposed protective factors appeared to be related (though not significantly) to increases in recidivism, rather than decreases. No DRAOR items demonstrated a protective effect for JI women.

Integrating Findings with Research

This dissertation addressed several important questions regarding the interaction of gender, race, risk assessment and recidivism using a sample of racially diverse JI women. While many findings are consistent with previous research, others are difficult to reconcile with existing literature. The following section discusses the results of this dissertation and attempts to situate them within the relevant bodies of literature described in the introductory chapters.

Race. Race remains a contentious and poorly understood issue within the correctional justice system (consider, for example, the case of *Ewert v Canada*, 2018[27]), and while results of this dissertation certainly reflect these concerns, they are unable to provide much clarification. For example, the composition of the sample used in this study reflects concerns about the over-representation of historically marginalized groups in correctional populations; despite the fact that Black individuals account for less than 4% of the adult population in Iowa (IDOC, 2017), 17.6% of the women in the sample were Black. Related to the issue of racial disproportionality are concerns regarding racial biases in the correctional system. As noted by Flores et al., (2016), concerns about racial bias permeate the discourse on risk assessment and sentencing practices. Though results of this dissertation cannot speak to bias in sentencing practices and none of the results can be unequivocally interpreted as proof of bias, certain findings, such as the lack of measurement invariance found across racial groups coupled with the generally improved prediction seen for White individuals relative to Black and Hispanic individuals do suggest that current assessment practices may not be suitable for JIPs belonging to historically marginalized groups. Scholars including Hannah-Moffat (2010), Holder (2014), and Van Eijk (2017) who have voiced strong concerns about the potential for risk assessments to exacerbate racial bias among marginalized populations would likely perceive this as evidence that the DRAOR should not be used with such persons.

[27] On June 13, 2018, the Supreme Court of Canada issues its decision in *Ewert v Canada,* in which the majority held that the Correctional Service of Canada breached its statutory duty to Jeffrey Ewert, a Metis inmate, when it used five actuarial risk assessment tools which had not been validated for use with Indigenous persons.

Given the extent of disproportionality seen in correctional systems, concerns about racial bias are legitimate. Also, as the focus of this dissertation was on risk-assessment following release, it is worth noting that although much of the criticism has been directed at the front-end of the process (i.e., differential arrest rates, the inclusion of risk assessment in sentencing) that the same concerns apply to post-conviction decision-making (Skeem & Lowenkamp, 2016). Equally, it is important to acknowledge that the disproportionality issue is born of multiple interacting causes. For example, Frase (2013) suggests that disparities in per capita incarceration rates reflect a number of factors, including: (a) differences in the prevalence, incidence, and nature of criminal behaviour committed by individuals of different races; (b) actual bias or discrimination; and (c) the unequal impact of intendedly race-neutral policies. Evidence from this dissertation speaks to the first factor. Black women were found to incur more violations and to return to prison for any reason significantly more often than their White and Hispanic counterparts. Black women also experienced these outcomes (i.e., failed) more quickly.

These results are consistent with those of Flores and colleagues' (2016), who also found that the Black defendants in their sample had a higher base rate of failure than White defendants (52% versus 39%). However, Flores et al. (2016) suggest that these differences were not indicative of bias and simply describe the behaviour of defendants in the criminal justice system. This conclusion is inconsistent with the perspective taken in this dissertation. While differences in the prevalence of criminal behaviour are relevant to understanding the over-representation of marginalized populations in correctional systems, the impact of systemic and longstanding racial prejudice in underlying societal structures on these base rates cannot be ignored.

Also, While Flores et al. (2016) found that failure rates for both White and Black defendants increased monotonically in accordance with risk categorizations on the Northpointe COMPAS (a community used actuarial assessment tool), that was not the case in this dissertation. Moreover, considerable differences were observed with respect predictive accuracy, or predictive fairness across race groups. For instance, as part of their study, Flores et al. (2016) also conducted a series of discrimination analyses comparing the relative predictive accuracy of COMPAS scores for Black and White defendants. Results indicated that the COMPAS strongly predicted recidivism for all defendants (*AUC* = .71), and that predictive accuracy was likewise high for both Black and White defendants, with *AUC* values of .70 and .69 respectively. Similarly, Skeem and Lowenkamp (2016) used the Post-Conviction Risk Assessment (PCRA; Johnson et al., 2011) to empirically examine the relationships among race, risk assessment, and recidivism in a sample of Black and White JIPs in federal custody. They concluded that PCRA scores strongly predicted both non-violent and violent rearrest (*AUC* = .71, .72 and .74, .75) for Black and White JIPs respectively, and subsequent regression analyses yielded a non-significant interaction term, suggesting that race did not moderate the predictive ability of the PCRA. Findings of similar analyses conducted in this dissertation were starkly at odds with these findings, with results demonstrating that predictive accuracy was significantly lower for Black JIPs compared to White JIPs of both genders. Possible explanations are discussed subsequently.

Gender. Justice involved women are widely recognized as one of the fastest growing correctional populations (e.g., Blanchette & Brown, 2006; Brown et al., 2017; Greiner et al., 2014; Guerino et al., 2011; Public Safety Canada, 2016) and have therefore

received increasing attention from correctional researchers and policymakers alike. As the majority of contemporary risk assessment tools were developed and validated using samples comprised predominantly of JI men (Chesney-Lind, 1997), a goal of this dissertation was to evaluate whether the DRAOR could be considered gender-neutral.[28] Results of this dissertation largely support the gender-neutrality of DRAOR as a risk assessment tool for White JIPs, as prediction was quite consistent across White men and women, regardless of outcome. That said, findings germane to the gender-neutrality/gender-salience debate are more difficult to interpret for Black and Hispanic JIPs, especially as the smaller sample size presented challenges when it came to the reliability of analyses for these groups. Nevertheless, results generally suggested that the predictive ability of DRAOR scores was more limited for Black and Hispanic women when compared to Black and Hispanic men.

A central objective of Study 3 was to explore the predictive ability of individual DRAOR items and to compare accuracy across genders. Contrary to expectations, most DRAOR items did not significantly predict recidivism, regardless of individuals' gender and race. Only five DRAOR items (substance abuse, interpersonal relationships, living situation, and high expectations) significantly predicted outcome, and importantly, only did so for Black and White men. Accordingly, gender differences were evident at this level of analysis, but they were somewhat difficult to reconcile with findings from not only women offender research and gender-difference focused research (e.g., Shaw & Hannah-Moffat, 2004; Van Voorhis, 2012) but also with much of the research conducted

[28] Brown, Serin, Forth, Bennell, Nunes, and Pozzulo (2017) use the term "gender neutral" to describe such assessment tools, which were developed based on gender-neutral theories of crimes, tested on largely male samples, and subsequently used on female offenders.

from the perspective of gender-neutrality (e.g., Andrews et al., 2012). For example, factors like substance use and personal/emotional factors that have been repeatedly found to be more salient for JI women (e.g., Cimino et al., 2015; Gobeil et al., 2016; Olver et al., 2014, etc.) were better predictors for men in this sample.

That said, although none of the 19 DRAOR items achieved statistical significance when predicting recidivism for women, several did demonstrate stronger associations with outcome for some women as compared to some men (i.e., variation was present across races and type of recidivism). Problem solving, employment, and prosocial identity were better predictors of technical violations and any returns for White women than they were for White men, and interpersonal factors (peer associations, anger, interpersonal relationships) predicted new offences more accurately for White women than White men. For Black women, only the attitudes toward authority item evidenced improved prediction when compared to Black men. No DRAOR items reliably predicted recidivism for Hispanic women, though attitudes toward authority, access to victims, and interpersonal relationships predicted technical violations and returns for Hispanic men. Thus, while results did not necessarily conflict with the literature on JI women, neither were they entirely consistent with expectations regarding purportedly gender-salient items.

In a 2017 report summarizing the results of contemporary research pertaining to women's risk and need factors, Brown indicates that there is now sufficient evidence to demonstrate that the broader domains of substance use and personal/emotional factors are women-salient predictors of recidivism. Though not borne out in this dissertation, her conclusions are grounded in a careful synthesis of other contemporary research

examining risk, need, and recidivism in JI women (e.g., Chang et al., 2015; Cimino et al., 2015; Gobeil et al., 2016; Kopak et al., 2015; Olver et al., 2004; etc.). Accordingly, results of this dissertation may be more exception than rule. Notably, in such a complex area of research, it is not surprising that inconsistent and sometimes contradictory findings emerge. For example, similar to the results seen in this dissertation, McCoy and Miller (2013) found that alcohol and drug problems predicted rearrests for the male contingent of their matched sample, but not for women.

Finally, the importance of nuance in women-focused correctional research cannot be overstated. The results of numerous studies (e.g., Brown & Motiuk, 2008; Yang et al., 2015) underscore the need to really delve deeper into higher order constructs like the Central Eight (Andrews & Bonta, 2010) and to consider other personal, emotional, historical, and circumstantial variables that might exert a mediating or moderating effect. For example, Brown and Motiuk (2008) discovered that age of onset and engaging in social drinking differentially impacted later substance use for young men and women. Similarly, Yang and colleagues (2015) found evidence of an interaction effect between gender, self-esteem, decision making confidence, and peer support, such that high self-esteem predicted rearrest for men and low self-esteem predicted rearrest for women.

Considered as a whole, the women-focused correctional research supports the notion of gender-neutrality over gender-salience or gender-specificity. The results of this dissertation are somewhat consistent with this view though it is also apparent that other factors such as the inclusion of a race variable and methodological limitations further complicate the interpretation of results related to the impact of gender. Given the lack of clear support for the utility of DRAOR with women found in this dissertation, the use of

gender-responsive adjunct approaches to risk assessment and case management (e.g., the Female Additional Manual [FAM]; de Vogel, de Vries Robbé, Van Kalmthout, & Place, 2014), is recommended.

Intersectionality. The intersectionality paradigm stipulates that a marginalized individual such as a Black or Hispanic woman would have a significantly different experience with the criminal justice system than would a White man or even a White woman, and the results of this dissertation are largely consistent with this perspective. Collectively, findings regarding the predictive ability of the DRAOR demonstrated that accuracy was significantly poorer for Black and Hispanic JIPs, and Black and Hispanic women in particular. Though correctional scholars have yet to come up with comprehensive, testable theoretical explanations for why non-White females engage in criminal behaviours, existing theories can be broadly categorized as falling into one of three theoretical positions: (a) the gender/race similarities hypothesis (i.e., the personal, interpersonal, and community-reinforcement (PIC-R) model and the developmental life-course perspective), (b) the gender/race difference hypothesis (i.e., the masculinity model and the racialized gender stereotype expectations model), and (c) the double-jeopardy effect.[29] Briefly, the gender/race similarities hypothesis argues that the pathways leading into involvement with the criminal justice system are similar regardless of gender and race, and the gender/race differences hypothesis argues the reverse. While this dissertation is not in a position to contribute to this debate, it does, however, have some bearing on the double-jeopardy effect. As described by Chesney-Lind (1996), Rafter

[29] Readers interested in a detailed discussion of how each of these models explains crime are encouraged to see Brown et al., (2014).

(1985), and Simpson (1989), the double-jeopardy effect asserts that Black girls and women are at the greatest disadvantage owing to this race/gender combination, and that they experience the most systemic discrimination. While this dissertation does not provide evidence of discrimination, it does provide evidence that the DRAOR may not adequately capture the risk and need factors most salient for Black women and that this group remains especially poorly understood in the community supervision setting.

This dissertation represents one of the first attempts to simultaneously examine the impact of race and gender. With a few notable exceptions (e.g., a correctional strategy designed specifically for Indigenous JI women in Canada), most research has focused on either race or gender, ignoring the multiplicative effects that this intersection of marginalized group membership engenders. Encouragingly, the indiscriminate use of assessment tools developed and validated on aggregate samples with JIPs belonging to marginalized subpopulations is increasingly recognized as problematic. This dissertation lends further credence to this argument, as results overwhelmingly suggested that the DRAOR, a case-planning tool that is used internationally, may not adequately capture the risk and need factors of particular relevance to these groups.

DRAOR research. Designed and developed in 2007 (Serin, 2007) as a structured professional judgment assessment tool to assist case managers, the DRAOR has been in use in Iowa and New Zealand for a decade. In this time, considerable research effort has been directed toward validating it with diverse JIP populations. Overall, results have supported its utility as a case planning tool, with initial pilot studies in both Iowa (Chadwick, 2014; Serin & Prell, 2012) and New Zealand (Tamatea & Wilson, 2009; Wilson & Tamatea, 2010) demonstrating that the DRAOR possessed acceptable

psychometric properties and predictive accuracy. Subsequent larger scale studies in New Zealand (i.e., Hanby, 2013; Lloyd, 2015) confirmed these findings and provided evidence of other desirable qualities such as the incremental predictive validity of the dynamic items comprising the DRAOR above and beyond that offered by purely static tools and the potential to improve prediction through reassessment. Since, numerous studies have supported the utility of the DRAOR with diverse JIP populations including male sex offenders (Averill, 2016; Smeth, 2013), high risk men (Yesberg & Polaschek, 2015), and youth JIPs (Fergusson, 2015; Muirhead, 2016). Overall, results of prior research examining the predictive ability of the DRAOR for the prediction of technical violations and overall returns to custody has yielded *AUC*s in the range of .60 - .70.

In view of the largely positive results seen in DRAOR research thus far, the results of this dissertation were surprising. While previous research examining the prediction of recidivism outcomes with low base rates (i.e., new offences; Chadwick, 2014 or sexual recidivism; Smeth, 2013) have reported much lower *AUC*s (i.e., .46 – .60), results in this range were observed for various outcomes in this study. As discussed previously, this dissertation is the first to disaggregate by race and gender simultaneously in an American sample, and the inclusion of Black and Hispanic individuals certainly influenced results. However, as the predictive accuracy observed for White men and women in the sample was also below expectations, it is recommended that future researchers seek to replicate these results in a similar but independent American sample.

Implications for Practice

The results, and more importantly, the methodological approach used in this dissertation are insufficient to inform theory. Nevertheless, findings from this research do

have a number of important implications for the day-to-day practice of correctional staff more broadly, and for those who use the DRAOR in particular. First, parole officers who routinely assess clients using the DRAOR should be mindful of how they use DRAOR scores to inform their supervision practices, especially with female clients and individuals belonging to historically marginalized groups. While the DRAOR has undeniable merit as a case planning tool, results of Study 1 demonstrate that the items comprising the DRAOR (and therefore also DRAOR scores) may not carry the same meaning for women of different races, or even from one assessment to the next for the same woman. As such, women's levels of risk and/or need should not be compared to others on the basis of their DRAOR scores. Likewise, if a woman's DRAOR scores do not change from one assessment to the next, it should not be assumed that her level of risk has not changed. Instead, case managers should consider individual items that are indicative of particular dynamic risks or strengths at each occasion, and seek to mitigate or maximize them, as appropriate. Moreover, given the accruing body of evidence demonstrating the predictive superiority of more proximal DRAOR assessments over distal ones, it is recommended that precedence be given to scores of the most recent DRAOR assessment.

Relatedly, results of Study 2 raise concerns about the utility of DRAOR subscale scores insofar as case planning and supervision practices are concerned. While DRAOR Total scores can provide a useful summary of risk level, subscale scores were not reliably associated with recidivism and were actually inversely related with recidivism rates for some women in this sample. Thus, in the absence of validated norms, parole officers should be careful not to use membership in a particular risk group or scores above a given threshold (on the DRAOR or any other assessment tool) as a justification for a particular

supervision approach. Notably, these results are applicable to all JIPs, not just JI women or persons belonging to historically marginalized groups. That said, recognition that risk of recidivism is perhaps even more nuanced for non-White women is required. To mitigate risk for these groups, parole officers should endeavor to tailor supervision strategies to address each woman's risks and needs while also considering how these approaches might interact with her immediate re-entry context.

Finally, the collective results of these studies and Study 3 in particular provide further evidence that assessment tools like the DRAOR which were developed and validated on aggregate samples are suitable for use with JIPs belonging to these majority groups (i.e., White JIPs). Minor exceptions aside, DRAOR scores reliably predicted recidivism for White men and women and consistently demonstrated improved predictive accuracy when compared to the results seen for Black and Hispanic JIPs. Importantly, for Black and Hispanic JIPs, DRAOR scores often significantly over- and under-predicted recidivism, which was not the case for White JIPs. Collectively, results also suggested that the DRAOR did not adequately capture salient risk and need factors for Black and Hispanic JIPs; it is recommended that consideration be given to possible adjunct approaches to address this shortcoming.

Limitations and Future Directions

Results of this dissertation should be considered in light of two types of limitations; (a) methodological issues, and (b) inadequate information about implementation. With respect to methodology, the first limitation that needs to be acknowledged is that this study relied on archival data. A primary drawback of archival data is that it limits the nature and scope of the research questions that can be examined.

For example, while the present doctoral research contributes to the field by focussing on JI women from historically marginalized groups, its ability to contribute to certain aspects of women offender theory such as an understanding of the pathways into crime was inherently limited. Previous research (e.g., Bloom et al., 2013; Chesney-Lind & Rodriguez, 1983; Salisbury & Van Voorhis, 2009, Van Voorhis, 2012, etc.) has highlighted the significance of factors such as childhood abuse, trauma, age of onset of substance use, and self-esteem in explaining female criminal behaviour, and it would have been informative to be able to factor this information into analyses. Likewise, more detailed information about important static risk factors like prior criminal history may have provided important context that was not available in the current research.

Relatedly, reliance on an archival dataset also imposed important limitations on sample size. While the overall sample size was quite large by the standards of most research on JI women, the number of Black and Hispanic JIPs in the sample was nevertheless insufficient given the nature of some of the analyses (i.e., factor analysis, calibration, item-level prediction) combined with the level of disaggregation (i.e., gender by race subsamples, prediction examined across levels of risk). Considered in conjunction with the characteristically low base rates of certain recidivism outcomes like new offences and the naturally smaller proportions of individuals with DRAOR scores falling at high and low ends of the possible range of scores, the number of Black and Hispanic individuals in the sample was inadequate.

The current study was also limited by how the outcome variables (i.e., any violation and new charges) were defined and provided by the host agency. As described in the limitations section of Study 2, the definitions for each type of recidivism likely

contributed to imprecision as they were quite broad. For example, technical violations were defined as any behaviour that violated one or more conditions of parole. Thus, violations ranged from minor (e.g., failing to abide by a curfew or missing a meeting with their parole officer) to serious (e.g., violating a no contact order, possessing or purchasing dangerous weapons, or failing to take prescribed psychotropic medication). This degree of variability may have significantly impacted base rates and predictive accuracy. Reliance on baseline assessment scores also presents a limitation; the current research cannot account for any reductions in recidivism that may have resulted from appropriate PO supervisions responses occurring after the first assessment was conducted.

With respect to implementation, no information about the supervision staff who completed the DRAOR assessments was available. Though all staff in Iowa receive training, other potentially influential staff variables (e.g., level of experience, training, belief in the utility of the DRAOR) could have impacted rating fidelity. Staff could also have differed in terms of how they responded to observed changes in individuals on their caseload (see Schaefer & Williamson, 2018; Viljoen et al., 2019).

Similarly, degree of "buy in" among supervision officers was unknown. As alluded to earlier, the level and amount of staff "buy in" can have a profound impact on the efficacy of an assessment measure (Latessa & Lovins, 2010; Schlagger, 2009). Staff members who are not convinced of the utility of a measure may be less inclined to complete it diligently. Miller and Maloney (2013) examined community corrections practitioners' compliance and non-compliance with risk and needs assessment tools. Staff who completed tools carefully and honestly tended to use them for decision making. However, other staff, characterized as "formal compliers", completed the tools to

conform to organizational expectations but filled them out in more perfunctory manner and often made decisions that did not correspond with tool results.

Finally, information pertaining to operational variables, such as training quality and compliance monitoring across IDOC jurisdictions was likewise unavailable. Although scoring manuals are available and standardized training and credentialing is a requirement, it may still be the case that there is variation among supervision officers in the use of the DRAOR. Other external factors (e.g., the quality and recency of their training) as well as internal factors (e.g., their level of conscientiousness and commitment to their job) could also impact scoring fidelity, which in turn would affect find findings regarding predictive validity.

In addition to attempting to address the above methodological limitations, future research should seek to advance the field's understanding of the correctional experiences of marginalized individuals by continuing to examine the influence of gender and race on risk assessment with JI women and historically marginalized groups while also continuing to explore how the DRAOR can contribute to this understanding. This doctoral research raised a number of important questions regarding the role of dynamic risk and protective factors with JI women, and continued research is needed to determine the extent to which these findings emerge in other studies. For example, in Study 2, women's recidivism rates were significantly associated with risk level, but recidivism did not always increase with risk as would be expected. Accordingly, there is a need to carefully examine the routine supervision practices employed with women, and with those assessed as high and low risk in particular. These findings may be an artefact of this particular sample, but if they are not, clarification of the underlying causes of this

observation would have important implications for case management. Likewise, it is recommended that future research examine the relationships between recidivism and items such as substance use and personal/emotional factors; these factors have increasingly been recognized as women-salient, but this was not the case in the present study.

Notably, this dissertation represents the first large-scale investigation of the DRAOR with a large sample of racially diverse women, and replication research is therefore inherently advised. The substandard findings regarding the predictive accuracy of DRAOR scores with White men underscore the need for further research as they suggest that irregularities may have been present in the data and add weight to concerns about scoring fidelity. Nevertheless, specific avenues of additional validation research on the DRAOR that merit attention include: (a) its factor structure, psychometric properties, and degree of measurement invariance with JI persons belonging to historically marginalized groups; (b) the functioning of DRAOR items from a test construction perspective (i.e., the proportion of overall test information provided by individual item); and (c) how the items in the Protective subscale function with different subgroups.

It is also recommended that the Iowa Department of Corrections endeavor to explore the scoring fidelity of parole officers when completing assessments. Specifically, examinations of (a) the quality and availability of DRAOR scoring training, (b) inter-rater reliability, (c) staff "buy-in" and compliance monitoring, (d) the impact of case manager – client relationships, and (e) how DRAOR scores are used to inform intervention strategies is recommended. Exploring whether jurisdictional differences exist might also be informative. Ultimately, a better understanding of factors affecting

scoring fidelity is required before firm conclusions can be made about the utility of the DRAOR with diverse correctional populations.

Conclusion

This dissertation is the first large-scale study to investigate the suitability of the DRAOR for use as a case management tool with racially diverse JI women and findings from the component studies collectively suggest that caution is required. Important limitations preclude firm conclusions regarding the appropriateness (or lack thereof) of the DRAOR for use with the subpopulations, but findings unequivocally indicate that further research is required to better understand the influence of gender and race on dynamic risk assessment and how supervision practices can be tailored to best meet the needs of these important correctional populations.

References

Adelman, J. (2020). *Somebody's watching me: examining the impact of probation officer caseloads on revocation rates* [Master's thesis, California State University]. Scholarworks. http://csus-dspace.calstate.edu/handle/10211.3/215072

Adler, F. (1975). *Sisters in crime: The rise of the new female criminal*. McGraw-Hill.

Allison, D. R. (1995). When is it worth measuring a covariate in a randomized clinical trial? *Journal of Consulting and Clinical Psychology*, *63*(3), 339-343. https://doi.org/10.1037/0022-006X.63.3.339

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

Andrews, D. (2012). *The Risk-Need-Responsivity (RNR) model of correctional assessment and treatment*. (J. Dvoskin, L. Skeem, R. Novaco, & K. Douglas Eds.). American Psychology-Law Society; Oxford University Press.

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

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

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI).* Toronto: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S., Guzzo, L., and Brews, A. (2008). *The relative predictive and incremental validity of gender-neutral and gender-informed risk/need* [Unpublished Manuscript]. Department of Psychology, Carleton University.

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, *56*(1), 113-133.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks.* Pro Publica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arluke, N.R. (1956). A summary of parole rules. *Crime & Delinquency, 2*(1), 6-13.

Arluke, N.R. (1969). A summary of parole rules – Thirteen years later. *Crime & Delinquency, 15*(2), 267-274.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397-438.

Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management, 41*(6), 1561-1577. https://journals.sagepub.com/doi/10.1177/0149206315591075

Averill, A. E. (2016). *An investigation of the Dynamic Risk Assessment for Offender Re-entry (DRAOR) with New Zealand sexual offenders* [Master's thesis, University of Canterbury]. University of Canterbury Research Repository. https://ir.canterbury.ac.nz/handle/10092/12883

Babchishin, K. M. (2013). *Sex offenders do change on risk-relevant propensities: Evidence from a longitudinal study of the Acute-2007* [Unpublished doctoral dissertation]. Carleton University.

Baca Zinn, M., & Thornton Dill, D. (1996). Theorizing differences from multiracial feminism. *Feminist Studies, 22*, 321–332.

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

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

Bandalos, B. (1996). *Applied multivariate statistics for the social sciences* (J. Stevens, Ed.). Lawrence Erlbaum Associates Inc.

Belknap, J. (2014). *The invisible woman: Gender, crime, and justice*. Nelson Education.

Benda, B. B. (2005). Gender differences in life-course theory of recidivism: A survival analysis. *International Journal of Offender Therapy and Comparative Criminology*, *49*(3), 325-342.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289-300.

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: receiver operating characteristic curves. *Critical Care*, *8*(6), 508.

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

Blanchette, K., Verbrugge, P., & Wichmann, C. G. (2002). *The Custody Rating Scale, initial security level placement, and women offenders*. Research Branch, Correctional Service of Canada.

Bloom, B., Owen, B., & Covington, S. (2002, November). *A theoretical basis for gender-responsive strategies in criminal justice* [Paper presentation]. 54th Annual Meeting of the American Society of Criminology, Chicago, IL, United States.

Bloom, B., Owen, B. A., & Covington, S. (2003). *Gender-responsive strategies: Research, practice, and guiding principles for women offenders*. National Institute of Corrections. https://www.ncjrs.gov/App/abstractdb/AbstractDBDetails.aspx?id=201301

Boer, D., Couture, J., Geddes, C., & Ritchie, A. (2003). *Yókw'tól: The understanding of one is complete - Risk management guide for Aboriginal offenders*. Aboriginal Initiatives Branch, Government of Canada.

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

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

Borum, R., Bartel, P. A., & Forth, A. E. (2005). Structured assessment of violence risk in youth. *Mental health screening and assessment in juvenile justice*, 311-323.

Brennan, T., Dieterich, W., & Oliver, W. (2007). *COMPAS: Correctional Offender Management for Alternative Sanctioning: Technical manual and psychometric report (V. 5.01).* Northpointe Institute for Public Management.

Britt, J. Y., Patton, C. L., Remaker, D. N., Prell, L., & Vitacco, M. J. (2019). Predicting violence risk and recidivism in female parolees: A state-wide sample. *International journal of law and psychiatry*, *66*, 101471.

Brook, J. S., Whiteman, M., Gordon, A. S., & Cohen, P. (1989). Changes in drug involvement: A longitudinal study of childhood and adolescent determinants. *Psychological Reports*, *65*(3), 707-726.

Brown, J. D. (2011). Likert items and scales of measurement. *Statistics*, *15*(1), 10-14.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Brown, S.L. (2017). *A review of the women offender risk/need research: In search of gender-neutral, women-salient and women-specific risk factors.* Research Branch, Correctional Service of Canada

Brown, S. L., Jones, N. J., & Greiner, L. (2014). Taking stock of the intersection of race, gender, and crime: Statistics, theory, and correctional applications. *Handbook of race-ethnicity and gender in psychology.* Springer.

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

Brown, S. L., Serin, R.C., Forth, A. E., Bennell, C., Nunes, K. L., & Pozzulo, J. (2017). *Psychology of criminal behaviour: A Canadian perspective* (Second ed.). Pearson Canada.

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

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111–150.

Byrne, M. K., & Howells, K. (2002). The psychological needs of women prisoners: Implications for rehabilitation and management. *Psychiatry, Psychology and Law*, *9*(1), 34-43.

Campbell, A. (1994). *The girls in the gang: A report from New York City.* Blackwell.

Campbell, A. (2002). *A mind of her own: The evolutionary psychology of women*. Oxford University Press.

Campbell, B. (2002). Multiracial feminism: Recasting the chronology of second wave feminism. *Feminist Studies, 28*, 337–360.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*(6), 567-590.

Canadian Institutes of Health Research (2019). Sex, Gender, and Health Research. https://cihr-irsc.gc.ca/e/50833.html

Carlen, P. (1988). *Women, crime and poverty*. Milton Keynes: Open University Press.

Carson, E. A. (2015). *Prisoners in 2014*. Bureau of Justice Statistics.

http://www.bjs.gov/index.cfm?ty=pbdetail&iid=5387

Caudy, M. S., Durso, J. M., & Taxman, F. S. (2013). How well do dynamic needs predict recidivism? Implications for risk assessment and risk reduction. *Journal of Criminal Justice, 41*, 458-466. doi:10.1016/j.jcrimjus.2013.08.004

Chadwick, N. (2014*). Validating the Dynamic Risk Assessment for Offender Re-entry (DRAOR) in a sample of U.S. probationers and parolees* [Master's thesis, Carleton University]. ProQuest Dissertations and Theses.

Chadwick, N., Dewolf, A., & Serin, R. (2015). Effectively training community supervision officers: A meta-analytic review of the impact on offender outcome. *Criminal Justice and Behavior*, *42*(10), 977-989.

Chanda, A. K., & Madalla, G. S. (1983). Erratum in: 1984. *Economics Letters*, *15*, 195-196.

Chang, Z., Larsson, H., Lichtenstein, P., & Fazel, S. (2015). Psychiatric disorders and violent reoffending: a national cohort study of convicted prisoners in Sweden. *The Lancet Psychiatry*, *2*(10), 891-900.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, *14*(3), 464-504.

Chesney-Lind, M. (1986). " Women and crime": The female offender. *Signs: Journal of Women in Culture and Society*, *12*(1), 78-96.

Chesney-Lind, M. (1996). *Race, gender, and class in criminology: The intersection* (M. D. Schwartz & D. Milovanovic,Eds.). Garland.

Chesney-Lind, M. (1997). *The female offender: Girls, women and crime***.** Sage Publications.

Chesney-Lind, M., & Pasko, L. (2013). *The female offender: Girls, women, and crime*. Sage Publications.

Chesney-Lind, M., & Rodriguez, N. (1983). Women under lock and key: A view from the inside. *The Prison Journal*, *63*(2), 47-65.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, *9*(2), 233-255.

Chilton, R., & Datesman, S. K. (1987). Gender, race, and crime: An analysis of urban arrest trends, 1960-1980. *Gender & Society*, *1*(2), 152-171.

Chu, C. M., Thomas, S. D., Ogloff, J. R., & Daffern, M. (2011). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health*, *10*(4), 337-345.

Cimino, A. N., Mendoza, N., Thieleman, K., Shively, R. & Kunz, K. (2015). Women reentering the community: Understanding addiction and trauma-related characteristics of recidivism. *Journal of Human Behavior in the Social Environment.* 25(5), 468-476.

Cloward, R. A., & Ohlin, L. E. (2013). *Delinquency and opportunity: A study of delinquent gangs*. Routledge.

Cochran, W. G. (1952). The $\chi 2$ test of goodness of fit. *The Annals of Mathematical Statistics*, 315-345.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 588-608.

Cohen, L. E., Felson, M., & Land, K. C. (1980). Property crime rates in the United States: A macrodynamic analysis, 1947-1977. *American Journal of Sociology*, *86*(1), 90-118.

Coid, J., Yang, M., Ullrich, S., Roberts, A., & Hare, R. D. (2009). Prevalence and correlates of psychopathic traits in the household population of Great Britain. *International Journal of Law and Psychiatry*, *32*(2), 65-73.

Cole, E. R. (1999). Intersectionality and research in psychology. *American Psychologist, 64*, 170–180.

Collins, P. H. (2001). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). Routledge.

Comrey, A. L., & Lee, H. B. (1992). *Interpretation and application of factor analytic results*. Hillsdale.

Cooper, A., & Smith, E. L. (2011). Homicide trends in the United States, 1980-2008 (NCJ 236018). *Bureau of Justice Statistics*.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-101.

Costa, F. M., Jessor, R., & Turbin, M. S. (1999). Transition into adolescent problem drinking: the role of psychosocial risk and protective factors. *Journal of Studies on Alcohol*, *60*(4), 480-490.

Costello, A., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, & Evaluation, 10*(7), 1-9.

Craig, L. A., & Beech, A. (2009). Best practice in conducting actuarial risk assessments with adult sexual offenders. *Journal of Sexual aggression*, *15*(2), 193-211.

Crenshaw, K.W. (1989/1993). *Feminist legal theory: Foundations* (D.K. Weisbert, Ed.). Temple University Press.

Daigle, L. E., Cullen, F. T., & Wright, J. P. (2007). Gender differences in the predictors of juvenile delinquency: Assessing the generality-specificity debate. *Youth Violence and Juvenile Justice*, *5*(3), 254-286.

Daly, M. (1992). Europe's poor women? Gender in research on poverty. *European Sociological Review*, *8*(1), 1-12.

Daly, K. (1994). *Gender, crime, and punishment*. Yale University Press.

Daly, K. (1998). *The handbook of crime and justice* (Tonry, Ed.). Oxford University Press.

Daly, K., & Chesney-Lind, M. (1988). Feminism and criminology. *Justice Quarterly, 5*(4), 101–143.

Daly, M., & Wilson, M. (1990). Killing the competition. *Human Nature*, *1*(1), 81-107.

Dell, C., Lyons, T., Grantham, S., Kilty, S., & Chase, W. (2014). *Researching with respect: The contributions of feminist, Aboriginal and community-based research approaches to the development of our study of First Nations women's healing from problematic drug use*. Innana Publications and Education Inc.

Derkzen, D., Harris, A., Wardrop, K., & Thompson, J. (2017). Outcomes of Women Offender Correctional Programs. *Advancing Corrections Journal*, *3*, 66-83.

Deschenes, E. P., Owen, B., & Crow, J. (2007). *Recidivism among female prisoners: Secondary analysis of the 1994 BJS recidivism data set* (Project No. 2004-IJ-CX-

0038) [Grant]. United States Department of Justice. https://www.researchgate.net/profile/Summer_Newell/publication/273352418_A_Comparative_Study_of_White_Asian_American_and_Other_non-White_men_and_Women_Under_Community_Supervision/links/564cdee808ae4988a7a3f963.pdf

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment*, *24*(30), 685-700. doi:10.1037/a0026668

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

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2012). SAPROF: *Structured assessment of protective factors for violence risk (2nd ed.).* Clinical Psychological Science.

de Vogel, V., de Vries Robbé, M., Van Kalmthout, W., & Place, C. (2014). Female Additional Manual (FAM). Additional guidelines to the HCR-20 V3 for assessing risk for violence in women. English version. *Utrecht, The Netherlands: Van der Hoeven Kliniek. Dolan, M. & Doyle, M.(2000). Violence risk prediction: Clinical and actuarial measures and the role of the Psychopathy Checklist. British Journal of Psychiatry*, *177*, 303-311.

de Vries Robbé, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *The Journal*

of Forensic Psychiatry & Psychology, *24*(4), 440–457. http://dx.doi.org/10.1080/14789949. 2013. 818162.

Dodge, M., & Progrebin, M. R. (2001). Collateral costs of imprisonment for women: Complications of reintegration. *Prison Journal, 81*(1), 42–54.

Doren, D. M. (2004). Stability of the interpretative risk percentages for the RRASOR and static-99. *Sexual Abuse: Journal of Research and Treatment, 16*(1), 25-36. doi:10.1177/107906320401600102

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

Dowden, C., & Andrews, D. A. (1999). *What works in young offender treatment: A meta-analysis* [Unpublished manuscript]. Department of Psychology, Carleton University.

Durrant, R. (2016). Putting risk factors in their place: an evolutionary-developmental approach to understanding risk. *Psychology, Crime & Law*, *22*(1-2), 17-32.

Edens, J. F., Campbell, J. S., & Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior*, *31*(1), 53-75.

Edin, K., & Kefalas, M. (2005). *Promises I can keep: Why poor women put motherhood before marriage.* University of California Press.

Embretson, S. E., & Steven, P. Reise. 2000. *Item response theory for psychologists.* Lawrence Erlbaum Assoc. Inc.

Ewert *v.* Canada, SCC 30 (2018). https://www.bccla.org/wp-content/uploads/2018/06/Ewert-en.pdf

Fabrigar, L. R., Wegener, D. T., MacCallum. R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. 10.1177/1094428104263675

Farrington, D. P. (2003). Developmental and life-course criminology: Key theoretical and empirical issues—The 2002 Sutherland award address. *Criminology*, *41*(2), 221-255. 10.1111/j.1745-9125.2003.tb00987.x

Farrington, D. P. (2007). *The Cambridge handbook of violent behavior and aggression* (D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman, Eds.). Cambridge University Press.

Ferguson, A. (2015). *An age-old issue: Evaluating the applicability of adult criminal risk assessment tools for use with youth offenders* [Master's thesis, Victoria University of Wellington]. Victoria University of Wellington, Creative Commons. http://researcharchive.vuw.ac.nz/handle/10063/4861

Field, A. (2000). *Discovering statistics using IBM SPSS statistics*. Sage Publications.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466.

Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to "machine bias: There's software used across the country to predict future criminals, and it's biased against blacks". *Federal Probation, 80*(2), 38-46.

Fortune, C. A., & Ward, T. (2017). Problems in protective factor research and practice. *Aggression and Violent Behavior*, *32*(1), 1-3. https://doi.org/10.1016/j.avb.2016.12.008

Fougere, A., & Daffern, M. (2011). Resilience in young offenders. *International Journal of Forensic Mental Health*, *10*(3), 244-253.

Frase, R. S. (2009). What explains persistent racial disproportionality in Minnesota's prison and jail populations? *Crime and Justice*, *38*(1), 201-280.

Frase, R. S. (2013). Research on race and sentencing: Goals, methods, and topics. *Justice Quarterly*, *30*(2), 262-269.

Gehring, K. S., & van Voorhis, P. (2014). Needs and pretrial failure: Additional risk factors for female and male pretrial defendants. *Criminal Justice and Behavior*, *41*(8), 943-970.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge accumulation. *Criminal Justice and Behavior*, *29*(4), 397-426.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, *34*(4), 401-433.

Geraghty, K.A., & Woodhams, J. (2015). The predictive validity of risk assessment tools for female offenders: A systematic review. *Aggression and Violent Behavior, 21*(1), 25-38.

Giordano, P. C., & Cernkovich, S. A. (1979). On complicating the relationship between liberation and delinquency. *Social Problems*, *26*(4), 467-481.

Gobeil, R., Blanchette, K., & Stewart, L. (2016). A meta-analytic review of correctional interventions for women offenders: Gender-neutral versus gender-informed approaches. *Criminal Justice and Behavior*, *43*(3), 301-322.

Gottfredson, S. D., & Moriatry, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, *52*(1), 178-200. doi:10.1177/0011128705281748

Greiner, L. E., Law, M. A., & Brown, S. L. (2014). Using dynamic factors to predict recidivism among women: A four-wave prospective study. *Criminal Justice and Behavior*, *42*(5), 457-480. 10.1177/0093854814553222

Green, L., & Campbell, M.A., (2006, June). *Gender influences and methodological considerations in adolescent risk-need assessment: A meta-analysis* [Paper presentation]. Canadian Psychological Association 67th Annual Meeting, Calgary, AB, Canada.

Greenfeld, L. A., & Snell, T. L. (1999). *Women offenders*. US Department of Justice, Bureau of Justice Statistics.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11), 78-94.

Guerino, P., Harrison, P. M., & Sabol, W. J. (2011). Prisoners in 2010. *Bureau of Justice Statistics*.

Guy-Sheftall, B. (1995). *Words of fire: An anthology of African-American feminist thought*. The New Press.

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.

Hanley, J. A. (1988). The robustness of the" binormal" assumptions used in fitting ROC curves. *Medical Decision Making*, *8*(3), 197-203.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Hanlon, T. E., Nurco, D. N., Bateman, R. W., & O'Grady, K. E. (1999). The relative effects of three approaches to the parole supervision of narcotic addicts and cocaine abusers. *The Prison Journal,* 79(2), 163-181.

Hannah-Moffat, K. (2004). Losing ground: Gendered knowledges, parole risk, and responsibility. *Social Politics: International Studies in Gender, State & Society*, *11*(3), 363-385.

Hannah-Moffat, K. (2010). Sacrosanct or flawed: Risk, accountability and gender-responsive penal politics. *Current Issues in Criminal Justice*, *22*(2), 193-215.

Hanson, R. K. (2017). Assessing the calibration of actuarial risk scales: A primer on the E/O index. *Criminal Justice and Behavior*, *44*(1), 26-39.

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*(1), 6-35. doi: 10.1177/0093854800027001002

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

Public Safety Canada. https://ccoso.org/sites/default/files/import/risk-assessment.pdf

Hanson, R. K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis.* Public Safety Canada. https://www.securitepublique.gc.ca/cnt/rsrcs/pblctns/ntmt-prtnr-vlnce/ntmt-prtnr-vlnce-eng.pdf

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

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human behavior*, *24*(1), 119-136.

Harcourt, B. E. (2015). Risk as a Proxy for Race. *Federal Sentencing Reporter*, *27*(4), 237-243.

Hare, R. D., & Frazelle, J. L. (1980). *Some preliminary notes on the use of a research scale for the assessment of psychopathy in criminal populations* [Unpublished manuscript]. Department of Psychology, University of British Columbia.

Harris, G. T., & Rice, M. E. (2015). Progress in violence risk assessment and communication: Hypothesis versus evidence. *Behavioral Sciences and the Law, 33*(1), 128–145.

Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the violence risk appraisal guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26*(4), 377-394.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003b). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*(3), 413-425. doi:10.1037/1040-3590.15.3.413

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: implications for substance abuse prevention. *Psychological Bulletin*, *112*(1), 64-73.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191-205.

Hazra, A. (2017). Using the confidence interval confidently. *Journal of Thoracic Disease*, *9*(10), 4125-4130.

Helmus, L. M., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*(1), 8-25. doi:10.1177/0093854816678898

Helmus, L., Hanson, R. K., & Thornton, D. (2009). Reporting Static-99 in light of new research on recidivism norms. *The Forum, 21*(1), 38-45.

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by static-99R and static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39*(9), 1148-1171. doi:10.1177/0093854812443648

Helmus, L., & Thornton, D. (2016). The MATS-1 risk assessment scale: Summary of methodological concerns and an empirical validation. *Sexual Abuse: Journal of Research and Treatment, 28*(3), 160-186. doi:10.1177/1079063214529801

Henderson, H., Tanana, M., Bourgeois, J. W., & Adams, A. T. (2015). Psychometric racial and ethnic predictive inequities. *Journal of Black Studies*, *46*(5), 462-481.

Hensher, D. A., & Johnson, L. W. (1981). Behavioural response and form of the representative component of the indirect utility function in travel choice models. *Regional Science and Urban Economics*, *11*(4), 559-572.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Hindelang, M. J., Gottfredson, M. R., & Garofalo, J. (1978). *Victims of personal crime: An empirical foundation for a theory of personal victimization*. Ballinger.

Hipwell, A. E., & Loeber, R. (2006). Do we know which interventions are effective for disruptive and delinquent girls? *Clinical Child and Family Psychology Review*, *9*(3), 221-255.

Holder, E. (2014, August 1). *Remarks as prepared for delivery* [Conference remarks]. Criminal Defense Lawyers 57th Annual Meeting, Philadelphia, PA, United States. https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153-166.

Holsinger, K. (2000). Feminist perspectives on female offending: Examining real girls' lives. *Women and Criminal Justice*, *12*(1), 23-52.

Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2006). Exploring the validity of the Level of Service Inventory-Revised with Native American offenders. *Journal of Criminal Justice*, *34*(3), 331-337.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117-144.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (1989). Logistic regression for matched case-control studies. *Applied Logistic Regression*, *2*(1), 223-259.

Howard, P. D., & Dixon, L. (2013). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys Violence Predictor. *Journal of Law and Human Behavior, 37*(3), 163-174. doi: 10.1037/lhb0000012

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1-55.

Hubbard, D. J., & Pratt, T. C. (2002). A meta-analysis of the predictors of delinquency among girls. *Journal of Offender Rehabilitation*, *34*(3), 1-13.

Huebner, B. M., DeJong, C., & Cobbina, J. (2010). Women coming home: Long-term patterns of recidivism. *Justice Quarterly*, *27*(2), 225-254.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581-592.

Iowa Department of Corrections (2016). *Iowa Department of Corrections FY2016 annual report.* https://doc.iowa.gov/sites/default/files/documents/2017/01/fy16_doc_annual_report_1.pdf

Iowa Department of Corrections (2017). *Quarterly quick facts: June 30, 2017.* https://doc.iowa.gov/data/quick-facts

Iowa Department of Corrections (2019). *Iowa Department of Corrections FY2019 annual report.*https://doc.iowa.gov/sites/default/files/documents/2019/11/fy2019_doc_annual_report.pdf

Jacobson, M. (2005). *Downsizing Prisons: How to reduce crime and end mass incarceration.* New York University Press.

Jessor, R., Van Den Bos, J., Vanderryn, J., Costa, F. M., & Turbin, M. S. (1995). Protective factors in adolescent problem behavior: Moderator effects and developmental change. *Developmental Psychology*, *31*(6), 923-933.

Johnson, J. L., Lowenkamp, C. T., VanBenschoten, S. W., & Robinson, C. R. (2011). The construction and validation of the federal Post Conviction Risk Assessment (PCRA). *Federal Probation*, *75*(1), 16-29.

Jones, N., Brown, S., Robinson, D., & Frey, D. (2015). Incorporating strengths into quantitative assessments of criminal risk for adult offenders the service planning instrument. *Criminal Justice and Behavior, 42*(3), 321-338. doi:10.1177/0093854814547041

Jones, N. J., Brown, S. L., & Zamble, E. (2010). Predicting criminal recidivism in adult male offenders: Researcher versus parole officer assessment of dynamic risk. *Criminal Justice and Behavior*, *37*(8), 860-882.

Kaeble, D., & Glaze, L.E. (2016). Correctional populations in the United States, 2015. *Bureau of Justice Statistics.*

Kim, E. S., Yoon, M., & Lee, T. (2012). Testing measurement invariance using MIMIC: Likelihood ratio test with a critical value adjustment. *Educational and Psychological Measurement*, *72*(3), 469-492.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). Guilford Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kopak, A. M., Proctor, S. L., & Hoffmann, N. G. (2015). Pathways to rearrest among court mandated female substance use treatment patients. *The American Journal on Addictions*, *24*(6), 495-498.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*(1), 337-343.

Kruttschnitt, C. (2001). *Women, Crime, and Criminal Justice* (C. Renzetti & L. Goodstein, Eds.). Roxbury Publishing.

Kruttschnitt, C., & Gartner, R. (2003). *Crime and justice: A review of research* (Tonry, Ed.). University of Chicago Press.

Kruttschnitt, C., & Gartner, R. (2005). *Marking time in the golden state: Women's imprisonment in California.* Cambridge University Press.

Labrish, C. S. (2011). *Advantages of using polychoric correlations for item-level exploratory factor analyses*. York University Press.

Latessa, E. J., Lemke, R., Makarios, M., & Smith, P. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). *Journal of Federal Probation*, *74*(1), 16-22.

Latessa, E. J., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims and Offenders*, *5*(3), 203-219.

Lewis, K., Olver, M. E., & Wong, S. C. (2012). The Violence Risk Scale: Predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment, 20*(2), 150-164. doi:10.1177/1073191112441242

Lloyd, C. D. (2015). *Can a dynamic risk instrument make short-term predictions in "real time"? Developing a framework for testing proximal assessment of offender recidivism risk during re-entry* (Unpublished doctoral dissertation). Carleton University,.

Lloyd, C., Hanson, K., Richards, D.K., & Serin, R.C. (2020). Reassessment improves prediction of criminal recidivism: A prospective study of 3,421 individuals in New Zealand. *Journal of Psychological Assessment, 32*(6), 568-581.

Lloyd, C.D., & Serin, R.C. (2012). Agency and outcome expectancies for crime desistance: Measuring offenders' personal beliefs about change. *Psychology, Crime, and Law*, 6, 543-565.

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.

Lodewijks, H. P. B., Ruiter, C. D., & Doreleijers, T. A. H. (2010). The impact of protective factors in desistance from violent reoffending: A study in three samples

of adolescent offenders. *Journal of Interpersonal Violence*, *25*(1), 568–587. http://dx.doi.org/10.1177/ 0886260509334403.

Lofthouse, R., Golding, L., Totsika, V., Hastings, R., & Lindsay, W. (2017). How effective are risk assessments/measures for predicting future aggressive behaviour in adults with intellectual disabilities (ID): a systematic review and meta-analysis. *Clinical Psychology Review*, *58*(4), 76-85.

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine*, *43*(2), 8–23.

Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2001). Risk/need assessment, offender classification, and the role of childhood abuse. *Criminal Justice and Behavior*, *28*(5), 543-563.

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*(1)*, 16-20.*

Luciani, F. P., Motiuk, L. L., & Nafekh, M. (1996). *An operational review of the custody rating scale: Reliability, validity and practical utility*. Research Division, Correctional Service of Canada.

Maher, L. (1997). *Sexed work: Gender, race and resistance in a Brooklyn drug market*. Clarendon.

Manchak, S. M., Skeem, J. L., Douglas, K. S., & Siranosian, M. (2009). Does gender moderate the predictive utility of the Level of Service Inventory—Revised (LSI-R) for serious violent offenders?. *Criminal Justice and Behavior*, *36*(5), 425-442.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320-341.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85-110.

Maruna, S. (2001). *Making good: How ex-convicts reform and rebuild their lives*. American Psychological Association.

Matsunaga, M. (2010). How to factor analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research, 3*(1), 97-110.

Matz, A. K., Conley, T. B., & Johanneson, N. (2018). What do supervision officers do? Adult probation/parole officer workloads in a rural Western state. *Journal of Crime and Justice*, *41*(3), 294-309.

Mauer, M. (1994). *Americans behind bars: The international use of incarceration, 1992-1993.* Sentencing Project, National Institute of Justice.

McCoy, L. A., & Miller, H. A. (2013). Comparing gender across risk and recidivism in nonviolent offenders. *Women & Criminal Justice*, *23*(2), 143-162.

McGrath, A., & Thompson, A. P. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Criminal Justice and Behavior*, *39*(3), 250-263.

McHugh M. L. (2013). The chi-square test of independence. *Biochemia medica*, *23*(2), 143-149. https://doi.org/10.11613/bm.2013.018

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, *7*(4), 361-388.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, *54*(1), 17-24.

Miller, H. A. (2006). A dynamic assessment of offender risk, needs, and strengths in a sam-ple of pre-release general offenders. *Behavioral Sciences & the Law, 24*(6), 767–782. http://dx.doi.org/10.1002/bsl.728.

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior*, *40*(7), 716-736.

Miller, J., & Mullins, C. (2006). *Taking stock: The status of criminological theory* (F.T. Cullen, J. Wright, & K. Blevins, Eds.). Transaction Publishers.

Mittlböck, M., & Schemper, M. (1996). Explained variation for logistic regression. *Statistics in Medicine*, *15*(19), 1987-1997.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, *13*(2), 355-375.

Morin, A. J. S., Arens, A., and Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of

construct-relevant psychometric multidimensionality. *Structural Equation Modeling 23*(1), 116–139. doi:10.1080/10705511.2014.961800

Morgan, R. D., Kroner, D. G., Mills, J. F., Serna, C., & McDonald, B. (2013). Dynamic risk assessment: A validation study. *Journal of Criminal Justice*, *41*(2), 115-124.

Muirhead, J. (2016). *Risky Business: Evaluating the Dynamic Risk Assessment for Offender Re-entry for Use with New Zealand Youth.* (Master's thesis, Victoria University of Wellington). Victoria University of Wellington Creative Commons. http://researcharchive.vuw.ac.nz/handle/10063/5221

Muthén, L. K., & Muthén, B. O. (2010). *Mplus software* (Version 6) [Computer Software]. Author.

Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D., & Martin, M. L. (2006). The Short-Term Assessment of Risk and Treatability (START) A prospective validation study in a forensic psychiatric sample. *Assessment*, *13*(3), 313-327.

Northepointe Inc. (1998). *Correctional Offender Management Profiling for Alternative Sanctions (COMPAS).* Author.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychological theory*. MacGraw-Hill.

Odgers, C. L., & Moretti, M. M. (2002). Aggressive and antisocial girls: Research update and challenges. *International Journal of Forensic Mental Health*, *1*(2), 103-119.

Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale-Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment*, *26*(5), 472-499. doi:10.1177/1079063213502679

Olver, M. E., Neumann, C. S., Wong, S. C., & Hare, R. D. (2013). The structural and predictive properties of the Psychopathy Checklist–Revised in Canadian Aboriginal and non-Aboriginal offenders. *Psychological Assessment*, *25*(1), 167.

Olver, M. E., & Sewall, L. A. (2018). Cross-validation of the discrimination and calibration properties of the VRAG-R in a treated sexual offender sample. *Criminal Justice and Behavior, 45*(6), 741-761. doi:10.1177/0093854818762483

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*, *36*(4), 329-353.

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, *26*(1), 156-176.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale – sexual offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment*, *19*(3), 318–329.

Orbis Partners. (2003). *Service Planning Instrument (SPIn)*. Author.

Orbis Partners. (2007a). *Youth Assessment and Screening Instrument: Brief Report* [Unpublished manuscript].

Orbis Partners. (2007b). *Youth Assessment and Screening Instrument: Girls (YASI-G).* Author.

Owen, B. (1998). *In the mix: Struggle and survival in a women's prison.* University of New York Press.

Owen, B. (2001). *Women, Crime and Criminal Justice* (C. Renzetti & L. Goodstein, Eds.). Roxbury.

Owen, B., & Bloom, B. (1995). Profiling women prisoners: Findings from the national surveys and the California sample. *Prison Journal, 75*(2), 165–185.

Pardoel, K. (2020). *Examining the shelf-life of baseline risk assessments with the Dynamic Risk Assessment for Offender Re-entry* [Unpublished manuscript]. Department of Psychology, Carleton University.

Perley-Robertson, B., Chadwick, N., & Serin, R.C. (2020). *Examining the calibration of the Dynamic Risk Assessment for Offender Re-entry* [Unpublished manuscript]. Department of Psychology, Carleton University.

Petersilia, J. (2001). Prisoner reentry: public safety and reintegration challenges. *The Prison Journal,* 81(3), 360-375.

Petersilia, J. (2003). *When prisoners come home: Parole and prisoner reentry.* Oxford Press.

Polaschek, D. L. L. (2016). Desistance and dynamic risk factors belong together. *Psychology, Crime & Law, 22*(1-2), 171-189. doi:10.1080/1068316X.2015.1114114

Pollak, O. (1950). *The criminality of women*. University of Pennsylvania Press.

Prell, L. (2013). *Iowa risk assessment revised: Predicting violence and victimization among male and female probationers and parolees* [Unpublished manuscript]. Iowa Department of Corrections.

Prell, L., Vitacco, M. J., & Zavodny, D. (2016). Predicting violence and recidivism in a large sample of males on probation or parole. *International Journal of Law and Psychiatry*, *49*(1), 107-113.

Public Safety Canada. (2015). *Corrections and conditional release statistical overview annual report*. http://www.publicsafety.gc.ca/res/cor/rep/2015-ccrso-eng.aspx

Public Safety Canada. (2016). *Corrections and conditional release statistical overview annual report*. http://www.publicsafety.gc.ca/res/cor/rep/2016-ccrso-eng.aspx

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*. American Psychological Association.

Rafter, N. H. (1985). *Partial justice: Women in state prisons, 1800-1935*. Northeastern University Press.

Raynor, P. (2007) Risk and need assessment in British probation: The contribution of the LSI-R. *Psychology, Crime, and Law*, *13*(2), 125-138.

Reynolds, C. R., & Livingston, R. B. (2013). *Mastering modern psychological testing: Theory & methods*. Pearson Higher Ed.

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*(5), 615-620. doi: 10.1007/s10979-005-6832-7

Richie, B. (2001). Challenges incarcerated women face as they return to their communities: Findings from life history interviews. *Crime and Delinquency, 47*(3), 368–389.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., Müller, M., Siegert, S., & Doering, M. (2020). *Package 'pROC': Display and analyze ROC curves* (Version 1.2) [Computer software]. Muthén & Muthén.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*(77)*,* 1-8. doi: 10.1186/1471-2105-12-77

Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*, *24*(5), 595-605. doi:10.1023/A:1005575113507

Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry, 147*(6), 598–611.

Salisbury, E. J., & Van Voorhis, P. (2009). Gendered pathways: A quantitative investigation of women probationers' paths to incarceration. *Criminal Justice and Behavior*, *36*(6), 541-566.

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

Scanlan, J., Fortune, C. A., Polaschek, D., & Yesberg, J. (2015, June 4-6). *Women on community supervision sentences: How do initial and proximal DRAOR scores compare in predicting recidivism* [Paper presentation]. North American Correctional and Criminal Justice Psychology Conference, Ottawa, ON, Canada.

Schaefer, L., & Williamson, H. (2018). Probation and parole officers' compliance with case management tools: Professional discretion and override. *International Journal of Offender Therapy and Comparative Criminology*, *62*(14), 4565-4584.

Schlager, M. D. (2018). Through the looking glass: Taking stock of offender reentry. *Journal of Contemporary Criminal Justice*, *34*(1), 69-80.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350-353.

Scott, T., (2017). *Risks, strengths, and recidivism among justice-involved youth: Investigating gender differences and similarities*. (Unpublished doctoral dissertation). Carleton University.

Seng, M., & Lurigio, A. J. (2005). Probation officers' views on supervising women probationers. *Women & Criminal Justice*, *16*(1), 65-85.

Serin, R.C. (2007). *The Dynamic Risk Assessment Scale for Offender Re-Entry (DRAOR): User Manual*. Author.

Serin, R.C., Chadwick, N., & Lloyd, C.D. (2015). Dynamic risk and protective factors. *Psychology, Crime & Law*, 22(1), 151-170.

Serin, R.C, Chadwick, N., & Pardoel, K. (2018). *Examining the Predictive Accuracy of the Dynamic Risk Assessment for Offender Reentry* [Unpublished manuscript]. Laura and John Arnold Foundation.

Serin, R. C., Gobeil, R., Lloyd, C. D., Chadwick, N., Wardrop, K., & Hanby, L. (2016). Using dynamic risk to enhance conditional release decisions in prisoners to improve their outcomes. *Behavioral Sciences & the Law, 34*(2-3), 321-336. doi:10.1002/bsl.2213

Serin, R. C., & Lloyd, C. D. (2009). Examining the process of offender change: The transition to crime desistance. *Psychology, Crime & Law*, *15*(4), 347-364.

Serin, R.C., Lloyd, C.L., & Chadwick, N. (2019). *Correctional Psychology* (D.L.L. Polaschek, A. Day, & C. Hollin, Eds.). Wiley International.

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

Serin, R. C., Mailloux, D., & Wilson, N. (2010). *Practice manual for use with Dynamic Risk Assessment for Offender Re-entry (DRAOR) scale: Revised version* [Unpublished manuscript]. Department of Psychology, Carleton University.

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

Shaw, M., & Hannah-Moffat, K. (2004). *What Matters in probation* (G. Mair, Ed.). Willan.

Silver, E., & Miller, L. L. (2002). A cautionary note on the use of actuarial risk assessment tools for social control. *Crime & Delinquency*, *48*(1), 138-161.

Simkins, S., & Katz, S. (2002). Criminalizing abused girls. *Violence Against Women*, *8*(12), 1474-1499.

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, *2*(1), 414-433.

Simourd, L., & Andrews, D. A. (1994, January). Correlates of delinquency: A look at gender differences. *Forum on Corrections Research, 6*(1), 26-31.

Simpson, S. S. (1989). Feminist theory, crime and justice. *Criminology, 27*(4), 605–631.

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

Singh, S. (2013). A dual problem of calibration design weights. *Statistics: A Journal of Theoretical and Applied Statistics, 47*(3), 566-574.

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

Skeem, J. (2013). Risk technology in sentencing: Testing the promises and perils (Commentary on Hannah-Moffat, 2011). *Justice Quarterly*, *30*(2), 297-303.

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: predictive bias and disparate impact. *Criminology*, *54*(4), 680-712.

Skeem, J.L., & Manchak, S. (2008). Back to the future: from Klockars' Model of effective supervision to evidence-based practice in probation. *Journal of Offender Rahabilitation, 47*(3), 220-247. doi:10.1080/10509670802134069

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, *20*(1), 38-42. doi: 10.1177/0963721410397271

Skeem, J., & Mulvey, E. (2002). *Care of the mentally disordered offender in the community* (A. Buchanan, Ed.). Oxford Press.

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research*, *102*(3), 443-461.

Smeth, A. (2013). *Evaluating risk assessments among sex offenders: a comparative analysis of static and dynamic factors* (Master's thesis, Carleton University). Carleton University CURVE. https://curve.carleton.ca/209c6f9c-c244-438a-ad60-80c72225e93a

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy*, *8*(1), 183-208.

Smith, D. A., & Visher, C. A. (1980). Sex and involvement in deviance/crime: A quantitative review of the empirical literature. *American Sociological Review*, *45*(4), 691-701.

Solomon, A.L., Jannetta, J., Elderbroom, B., Winterfield, L., Osborne, J., Burke, P., Stroker, R.P., Rhine, E.E., & Burrell, W.D. (2008). *Putting public safety first: 13 strategies for successful supervision and reentry* [Policy brief]. Urban Institute.

Sorbello, L., Eccleston, L., Ward, T., & Jones, R. (2002). Treatment needs of female offenders: A review. *Australian Psychologist*, *37*(3), 198-205.

Starr, S. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review, 66*(1), 803–872.

Statistics Canada. (2016). *Police-reported crime statistics in Canada, 2016.* http://www.statcan.gc.ca/pub/11-627-m/11-627-m2017031-eng.htm

Statistics Canada. (2017). Definitions, data sources and methods. https://www.statcan.gc.ca/eng/concepts/index

Steffensmeier, D. J., & Allan, E. A. (1988). Sex disparities in arrests by residence, race, and age: An assessment of the gender convergence/crime hypothesis. *Justice Quarterly*, *5*(1), 53-80.

Steffensmeier, D., & Allan, E. (1996). Gender and crime: Toward a gendered theory of female offending. *Annual Review of Sociology*, *22*(1), 459-487.

Stockdale, K. C., Olver, M. E., & Wong, S. C. (2010). The Psychopathy Checklist: Youth Version and adolescent and adult recidivism: considerations with respect to gender, ethnicity, and age. *Psychological assessment*, *22*(4), 768-781.

Stuart, B., & Brice-Baker, J. (2004). Correlates of higher rates of recidivism in female prisoners: An exploratory study. *Journal of Psychiatry and the Law, 32*(1), 29–70.

Sullivan, E., Mino, M., Nelson, J., & Pope, J. (2002). *Families as a resource in recovery from drug abuse: An evaluation of la bodega de la familia.* The Vera Institute of Justice.

Sutherland, E., & Cressey, D. (1978). *Criminology*. JP Lippincott.

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

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

Thompson, B. (2002). Multiracial feminism: Recasting the chronology of second wave feminism. *Feminist Studies, 28*(2), 337–359.

Tóth-Király, I., Bõthe, B., Rigó, A., & Orosz, G. (2017). An illustration of the exploratory structural equation modeling (ESEM) framework on the passion scale. *Frontiers in Psychology*, *8*(1), 1968-1980.

Travis, L.F., & Latessa, E.J. (1984). A summary of parole rules – thirteen years later: revisited thirteen years later. *Journal of Criminal Justice, 12*(6), 591-600.

Travis, L.F., & Stacey, J. (2010). A half century of parole rules: Conditions of parole in the United States, 2008. *Journal of Criminal Justice, 38*(4), 604-608. doi: 10.1016/j.jcrimjus.2010.04.032

Trevethan, S., Moore, J. P., & Rastin, C. J. (2002). *A profile of Aboriginal offenders in federal facilities and serving time in the community.* Research Branch, Correctional Service of Canada. https://www.researchgate.net/profile/Christopher_Rastin/publication/322209097_A_profile_of_Aboriginal_offenders_in_federal_facilities_and_serving_time_in_the_community/links/5a4b9fbe0f7e9b8284c1d6dd/A-profile-of-Aboriginal-offenders-in-federal-facilities-and-serving-time-in-the-community.pdf

Tsigilis, N., Gregoriadis, A., Grammatikopoulos, V., & Zachopoulou, E. (2018). Applying exploratory structural equation modeling to examine the Student-Teacher Relationship Scale in a representative Greek sample. *Frontiers in Psychology*, *9*(1), 733-777.

Ullrich, S., & Coid, J. W. (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 (doi:10/1037/a0023613).

Uanhoro, J. O., Wang, Y., & O'Connell, A. A. (2019). Problems with using odds ratios as effect sizes in binary logistic regression and alternative approaches. *The Journal of Experimental Education*, 1-20. https://doi.org/10.1080/00220973.2019.1693328

Van Dam, N. T., Earleywine, M., & Borders, A. (2010). Measuring mindfulness? An item response theory analysis of the Mindful Attention Awareness Scale. *Personality and Individual Differences*, *49*(7), 805-810.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

van Eijk, G. (2017). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. *Punishment & Society, 19*(4), 463-481. doi:10.1177/1462474516666282

Van Voorhis, P. (2012). On behalf of women offenders. *Criminology & Public Policy*, *11*(2), 111-145.

Van Voorhis, P., Bauman, A., & Brushett, R. (2013). *Revalidation of the Women's Risk Needs Assessment: Probation results.* United States National Institute of Corrections. https://static1.squarespace.com/static/5616ba8fe4b051d6d43c985b/t/57745ce36a4963b824f16048/1467243749602/Probation+Final+Report+2013+WEB.pdfVan

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

https://www.unlv.edu/sites/default/files/page_files/27/WRNA-Final-Report-to-National-Institute-of-Corrections-2008.pdf

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, *37*(3), 261-288.

Viljoen, J. L., Mordell, S., & Beneteau, J. L. (2012). Prediction of adolescent sexual reoffending: A meta-analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99. *Law and Human Behavior*, *36*(5), 423.

Viljoen, J. L., Shaffer, C. S., Muir, N. M., Cochrane, D. M., & Brodersen, E. M. (2019). Improving case plans and interventions for adolescents on probation: The implementation of the SAVRY and a structured case planning form. *Criminal Justice and Behavior*, *46*(1), 42-62.

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

Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior, 40*(12), 1383-1396. doi:10.1177/0093854813508916

Wagner, P. & Walsh, A. (2016, June 16). *States of incarceration: The global context 2016* [Press release]. http://www.prisonpolicy.org/global/2016.html

Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, *29*(4), 364-376.

Ward, T. (2017). Prediction and agency: The role of protective factors in correctional rehabilitation and desistance. *Aggression and Violent Behavior, 32*(1), 19-28. doi:10.1016/j.avb.2016.11.012

Ward, T., & McDonald, I. (2016). *Beyond the Risk Paradigm in Criminal Justice* (C. Trotter, G. McIvor, & F. McNeil, Eds.). MacMillan Education.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *The Short-Term Assessment of Risk and Treatability (START)(Version 1.1): User Manual.*

Webster, C. D., Martin, M., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *The Short-Term Assessment of Risk and Treatability (START) (Version 1): User Manual, Consultation Edition.*

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Consulting and Clinical Psychology*, *61*(6), 952-965. doi:10.1037/0022-006X.61.6.952

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, *10*(2), 395-426.

Wilson, C. M, Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior, 37*(6), 377-388.

Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*(5), 339-361. 10.1177/0146621611405984

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1-26.

Xu, R., & O'Quigley, J. (1999). A R^2 type measure of dependence for proportional hazards models. *Journal of Nonparametric Statistics*, *12*(1), 83-107. doi:10.1080/10485259908832799

Yang, Y., Knight, K., Joe, G.W., Rowan, G.A., Lehman, W.E.K., Flynn, P.M. (2015). Gender as a moderator in predicting re-arrest among treated drug-involved offenders. *Journal of Substance Abuse Treatment, 49*(1)*,* 65-70.

Yang, S., & Mulvey, E. P. (2012). Violence risk: Re-defining variables from the first-person perspective. *Aggression and Violent Behavior*, *17*(3), 198-207.

Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, *136*(5), 740-767.

Yesberg, J. A., & Polaschek, D. L. L. (2015). Assessing dynamic risk and protective factors in the community: Examining the validity of the Dynamic Risk

Assessment for Offender Re-entry. *Psychology, Crime & Law, 21*(1), 80-99. doi:10.1080/1068316X.2014.935775

Yesberg, J. A., Scanlan, J. M., & Polaschek, D. L. (2014). Women on parole: Do they need their own DRAOR? *The New Zealand Corrections Journal*, *2*(1), 20-25.

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

Yoshioka, M. R., & Noguchi, E. (2009). The developmental life course perspective: A conceptual and organizing framework for human behavior and the social environment. *Journal of Human Behavior in the Social Environment*, *19*(7), 873-884.

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

Appendix A: Certification of Institutional Ethics Clearance

Carleton UNIVERSITY

Office of Research Ethics
503 Robertson Hall | 1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 4085
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-B (CUREB-B) at Carleton University has renewed ethics clearance for the research project detailed below. CUREB-B is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

Title: Preliminary validation of the Dynamic Risk Assessment for Offender re-entry (DRAOR)

Protocol #: 106316 14-073

Principal Investigator: Ralph Serin

Department and Institution: Faculty of Arts and Social Sciences\Psychology (Department of), Carleton University

Project Team (and Roles): Ralph Serin (Primary Investigator)
Stephanie Biro (ACVS Staff)
Nick Chadwick (Research Assistant)
Kaitlin Pardoel (Research Assistant)
Kaitlyn Wardrop (Research Assistant)
Bronwen Perley-Roberts (Research Assistant)
Daniela Corno (Research Assistant)
Thea Froehlich (Research Assistant)
Taylor Carty (Research Assistant)

Funding Source (If applicable):

Effective: **December 16, 2019** Expires: **December 31, 2020**.

Please ensure the study clearance number is prominently placed in all recruitment and consent materials: CUREB-B Clearance # 106316.

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.

2. Any modification to the approved research must be submitted to CUREB-B. All changes must be approved prior to the continuance of the research.

3. An Annual Application for the renewal of ethics clearance must be submitted and cleared by the above date. Failure to submit the Annual Status Report will result in the closure of the file. If funding is associated, funds will be frozen.

4. A closure request must be sent to CUREB-B when the research is complete or terminated.

5.During the course of the study, if you encounter an adverse event, material incidental finding, protocol deviation or other unanticipated problem, you must complete and submit a Report of Adverse Events and Unanticipated Problems Form, found here:https://carleton.ca/researchethics/forms-and-templates/

6. It is the responsibility of the student to notify their supervisor of any adverse events, changes to their application, or requests to renew/close the protocol.

7. Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2nd edition* and the *Carleton University Policies and Procedures for the Ethical Conduct of Research* may result in the suspension or termination of the research project.

Upon reasonable request, it is the policy of CUREB, for cleared protocols, to release the name of the PI, the title of the project, and the date of clearance and any renewal(s).

Please email the Research Compliance Coordinators at ethics@carleton.ca if you have any questions.

CLEARED BY: **Date: December 16, 2019**

Natasha Artemeva, PhD, Chair, CUREB-B

Janet Mantler, PhD, Vice-Chair, CUREB-B

Appendix B: Measures Used in Dissertation

B1. The Dynamic Risk Assessment for Offender Re-Entry (DRAOR)

Stable Risk Indicators Characteristics associated with risk capable of changing over months or years				
Indicator	**Scoring Criteria**	**Score** (omit if unknown)		
Peer Associations	Has only prosocial peers (0) – Has only antisocial peers (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Attitudes Toward Authority	Prosocial attitudes (0) – Antagonistic attitudes (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Impulse Control	Autonomous/self monitoring (0) – Highly Impulsive (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Problem Solving	Ability to make good decisions (0) – No consideration of consequences (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Sense of Entitlement	Recognition of their limitations (0) – Inflated sense of self-worth (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Attachment with Others	Connected/concerned about others (0) – Callous/indifferent towards others (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Total STABLE Risk score /12				

Acute Risk Indicators Characteristics associated with risk capable of changing in the short term (<1 month).				
Indicator	**Scoring Criteria**	**Score** (omit if unknown)		
Substance Abuse	Maintaining sobriety/social use (0) – Problematic substance abuse (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Anger/Hostility	Absence of anger/hostility (0) – Marked presence of anger/hostility (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Opportunity/Access to Victims	Avoidance of preferred victims (0) – Access to preferred victims (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Negative Mood	No evidence of depression/anxiety (0) – Marked presence of depression/anxiety (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Employment	Maintaining a job (0) – Unemployed (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem

Interpersonal Relationships	In stable healthy relationship (0) – No relationship/conflicted relationship (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Living Situation	Stable and positive living situation (0) – Instability/lack of accommodations (2)	0 Not a problem	1 Slight/Possible Problem	2 Definite Problem
Total ACUTE Risk score /14				

Protective Factors Characteristics that may buffer risk.				
Indicator	**Scoring Criteria**	**Score** (omit if unknown)		
Responsive to Advice	Follows direction from prosocial peers, partners, supervisor, etc.	0 Not an asset	1 Slight/Possible asset	2 Definite asset
Prosocial Identity	Legitimately views self as no longer criminally oriented with behavioural examples.	0 Not an asset	1 Slight/Possible asset	2 Definite asset
High Expectations	Individual, family, and/or community have high expectations of success.	0 Not an asset	1 Slight/Possible asset	2 Definite asset
Costs/Benefits	Evidence that rewards of prosocial behaviour outweigh those of procriminal behaviour.	0 Not an asset	1 Slight/Possible asset	2 Definite asset
Social Support	Evidence that meaningful and accessible prosocial supports exist.	0 Not an asset	1 Slight/Possible asset	2 Definite asset
Social control	Conformity and compliance with prosocial others; Strong internalized connection/bond.	0 Not an asset	1 Slight/Possible asset	2 Definite asset
Total PROTECTIVE Factors score /12				
Total DRAOR score /26 **(Stable + Acute) - Protective**				

Risk Level	**DRAOR Cut-off**
Low	< 2
Moderate	3 to 9
Moderate/High	10 to22
High	≥ 23

B2. Iowa Violence and Victimization Instrument (IVVI)

	Violence Score	Victimization Score
Current Offense		
1. Active Offenses include…		
a. Assault, attempted murder, burglary, robbery, murder, theft from a person, vandalism or voluntary manslaughter	2	2
b. Not as above but most serious offense is forgery/fraud	-1	0
c. Not as above	0	0
2. Number of counts, current property offenses		
a. None	0	0
b. One	0	1
c. Two or more	0	2
Criminal History – Volume and Seriousness		
3. Ever convicted of murder/manslaughter, robbery or theft from a person (priors only).		
a. Yes	1	1
b. No	0	0
For items 4 & 5, consider only convictions where the date of conviction or sentencing is 10 years or less from the earliest sentence date of the current offense(s).		
4. Number of prior counts for violent crimes within last 10 years (any offense level).		
a. None	0	0
b. One to three	1	1
c. Four or more	2	2
5. Prior convictions within the last 10 years (check ALL that apply). *For a through c, count only aggravated misdemeanors, felonies, or juvenile commitment offenses:*	0	1
a. Property crime	1	1
b. Burglary (include violent and property offense type)	1	1
c. Weapons	1	1
d. Flight/escape (any offense level)	0	0
e. Not as above		
Criminal History – Recency		
6. Prior conviction for violent crime within the last 5 years (any offense level). *Consider only convictions where the date of conviction or sentencing is 5 years or less from the earliest sentence date of the current offense.*	2	2
a. Yes	0	0
b. No		
7. Released from Prison or Juvenile Commitment in the last 5 years for (check all that apply). *Count from last release date to current prison admission date.*	2	2
a. Violent crimes	0	1
b. Property crime	0	0
c. Not as above		
Criminal Orientation/Associates		
8. Security threat group membership:		
a. Confirmed member	3	3
b. Suspected or none	0	0

Current Age		
9. Current Age		
a. 24 or younger	2	1
b. 25-29	2	0
c. 30-37	1	0
d. 38-54	0	0
e. 55+	0	-1

Victimization scores are calculated first, and then are adjusted upwards or downwards following consideration of the violence score.

Victimization Score categories		**Violence Score Categories**	
Low	-1 to 2	Low	-1 to 2
Low/Moderate	2 to 3	Moderate	3 to 5
Moderate/High	4 to 7	High	6 to 9
High	8+	Very High	10+

Appendix C: Demographic Information for 500 Women EFA Subsample

Table C1

Demographic Information for Women in Randomly Selected Subsample of 500 for EFA

	All Women (*N* = 500)	White (*N* = 380)	Black (*N* = 76)	Hispanic (*N* = 26)
Age				
18-20	10	6	†	†
21-30	191	135	40	16
31-40	175	145	20	10
41-50	78	58	7	†
51+	46	37	8	†
Marital Status				
Single	253	184	53	16
Common-law	†	†	†	†
Married	81	68	7	†
Divorced	69	61	5	†
Separated	26	26	†	†
Widowed	8	7	†	†
Highest Education				
Grade 8 or lower	12	10	†	†
Some high school	70	56	7	†
High school diploma/GED	343	272	50	20
Post-secondary or other vocational training	45	34	†	†
Unknown	30	23	7	†

Note. † = cells with *N* of 5 or less are supressed.

Appendix D: Cross-Loadings for Alternative ESEM Models

Table D1

Factor Cross-loadings for Alternative Three-Factor Exploratory Structural Equation Model

	Risk I	Risk II	Protective
Peer Associations	.018	**.644**	-.125
Attitudes to Authority	**.668**	.002	-.231
Impulse Control	.391	**.644**	.033
Problem Solving	.376	**.592**	-.032
Sense of Entitlement	**.661**	-.038	-.138
Attachment with Others	**.400**	.151	-.205
Substance Abuse	-.009	**.528**	-.061
Anger / Hostility	**.612**	.053	.014
Access to Victims	**.262**	.152	-.023
Negative Mood	.178	**.465**	.044
Employment	-.010	**.414**	-.107
Interpersonal Relationships	-.020	**.535**	-.060
Living Situation	-.056	**.533**	-.130
Responsiveness to Advice	-.241	.007	**.648**
Prosocial Identity	-.029	-.140	**.675**
High Expectations	.006	-.02	**.758**
Costs / Benefits	-.137	-.057	**.684**
Social Support	.120	.004	**.769**
Social Control	.114	.007	**.903**

Note. Grey loadings were non-significant, and **bolded** loadings represent the alternative 3-factor structure.

Table D2

Factor Cross-Loadings for Four-Factor Exploratory Structural Equation Model

	Risk I	Risk II	Risk III	Protective
Peer Associations	**.735**	-.062	-.059	-.058
Attitudes to Authority	.018	**.679**	.049	-.184
Impulse Control	**.692**	.083	.377	.018
Problem Solving	**.635**	.087	.353	-.050
Sense of Entitlement	-.030	**.627**	.136	-.112
Attachment with Others	.180	**.447**	-.082	-.143
Substance Abuse	**.596**	-.198	.156	-.054
Anger / Hostility	.070	**.687**	-.044	.093
Access to Victims	.174	**.228**	.044	.001
Negative Mood	**.527**	.095	.027	.092
Employment	**.466**	-.066	-.035	-.071
Interpersonal Relationships	.634	.029	-.274	.047
Living Situation	.629	-.011	-.280	-.030
Responsiveness to Advice	.004	-.208	-.070	**.648**
Prosocial Identity	-.164	.054	-.063	**.684**
High Expectations	-.028	.005	.039	**.758**
Costs / Benefits	.040	-.005	-.194	**.743**
Social Support	.006	-.058	.355	**.757**
Social Control	.001	-.007	.244	**.870**

Note. Grey loadings were non-significant, and **bolded** loadings represent the alternative 3-factor structure.

Appendix E: Log Minus Log Plots of Hazard Functions

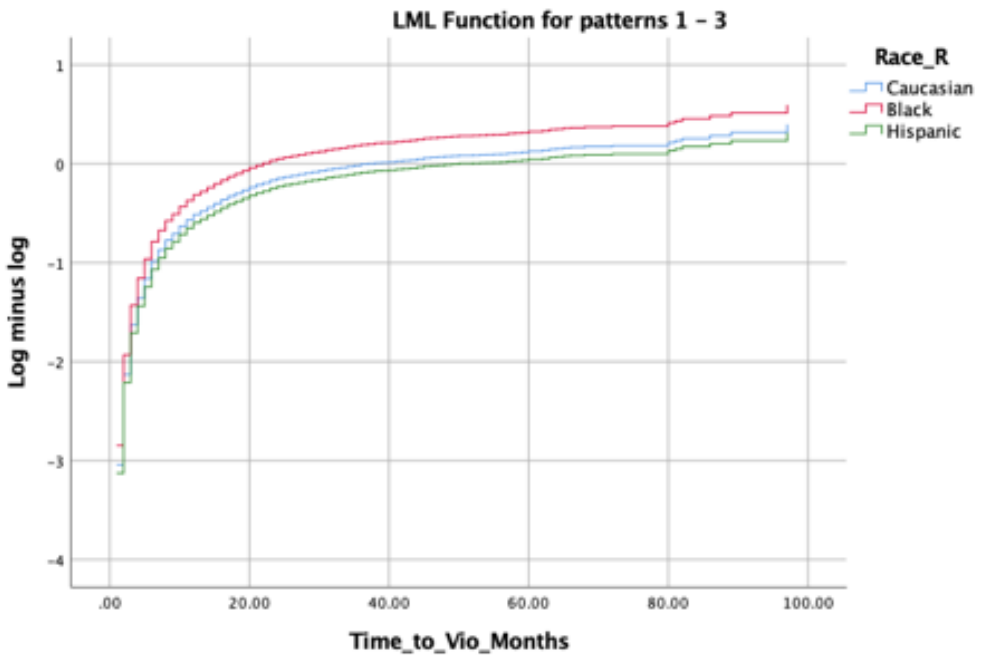


Figure E1. *Technical Violations by Race.*

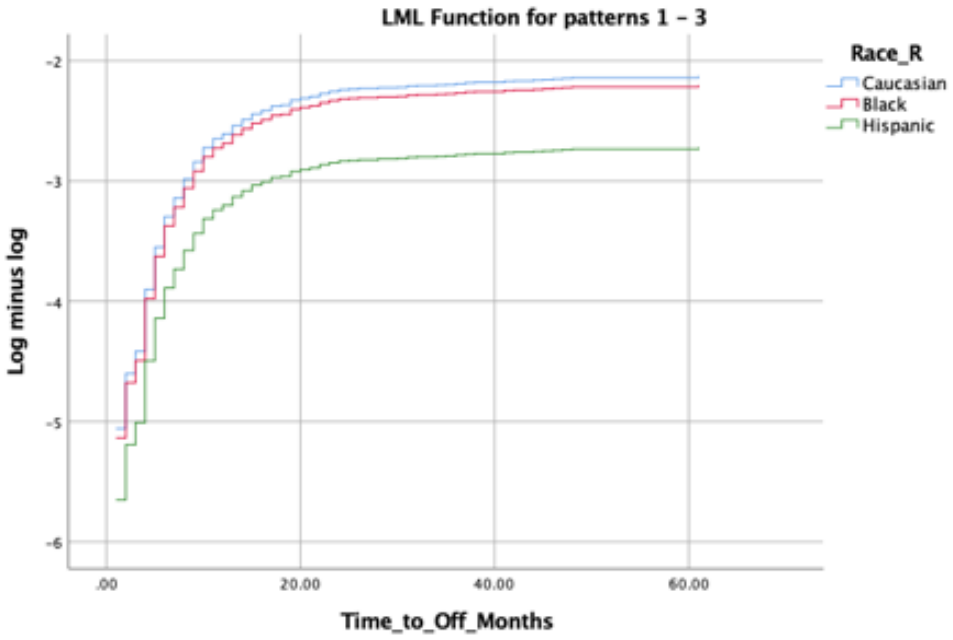


Figure E2. *New Offence by Race.*

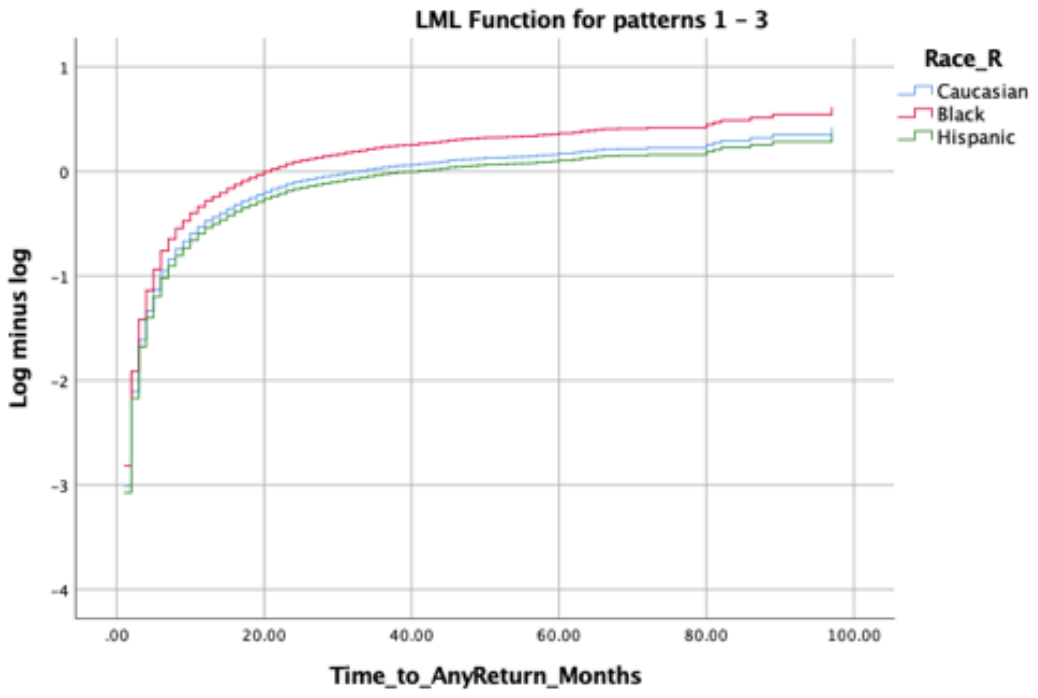


Figure E3. *Any Return by Race.*

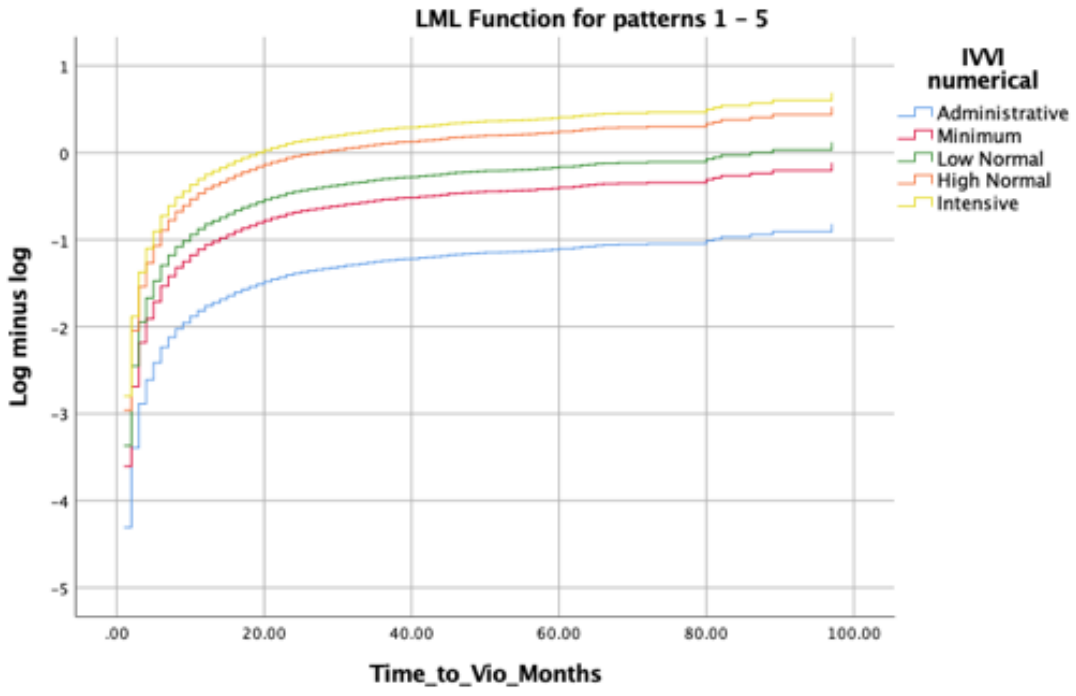


Figure E4. *Technical Violation by Level of Supervision as Determined by the IVVI.*

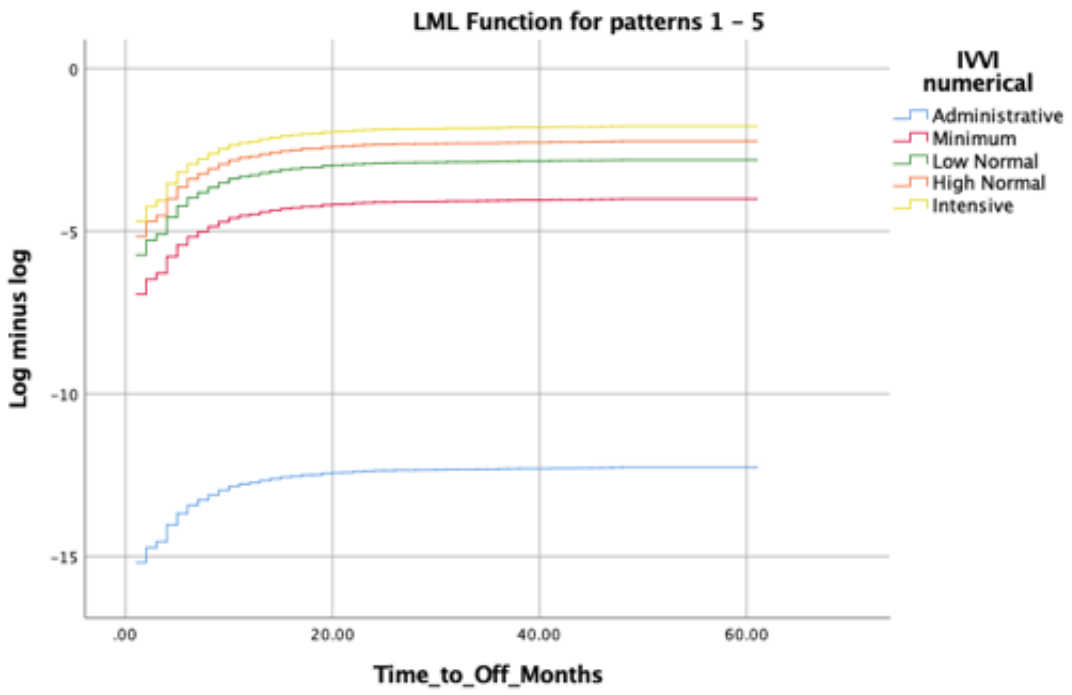


Figure E5. *New Offence by Level of Supervision as Determined by the IVVI.*

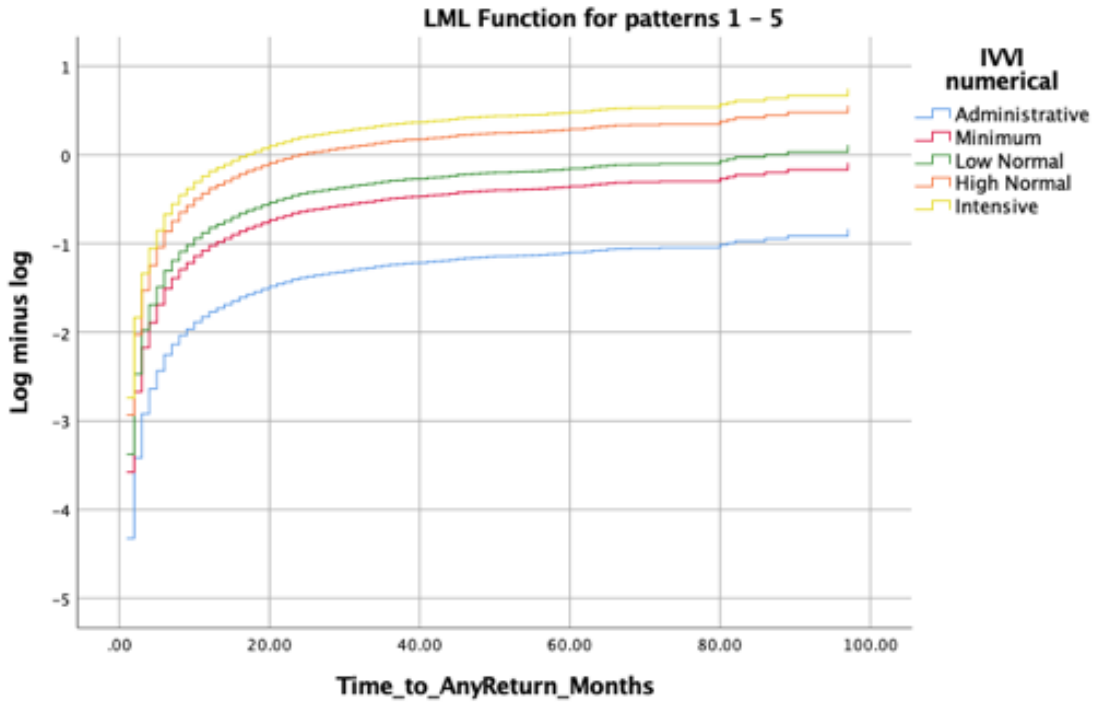


Figure E6. *Any Return by Level of Supervision as Determined by the IVVI.*

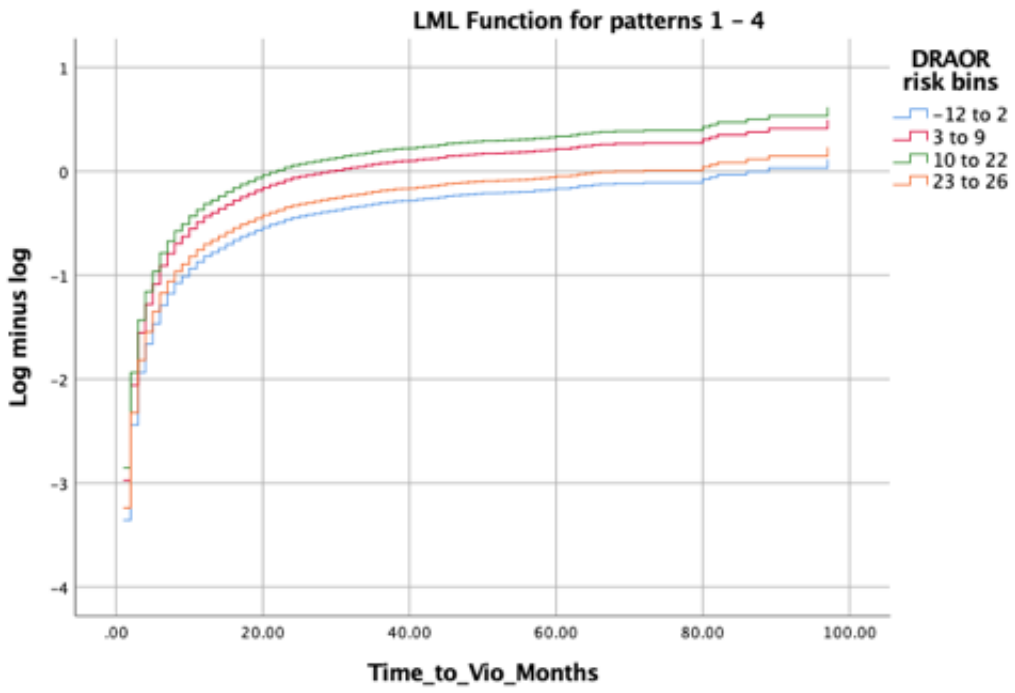


Figure E7. *Technical Violation by DRAOR Total Score.*

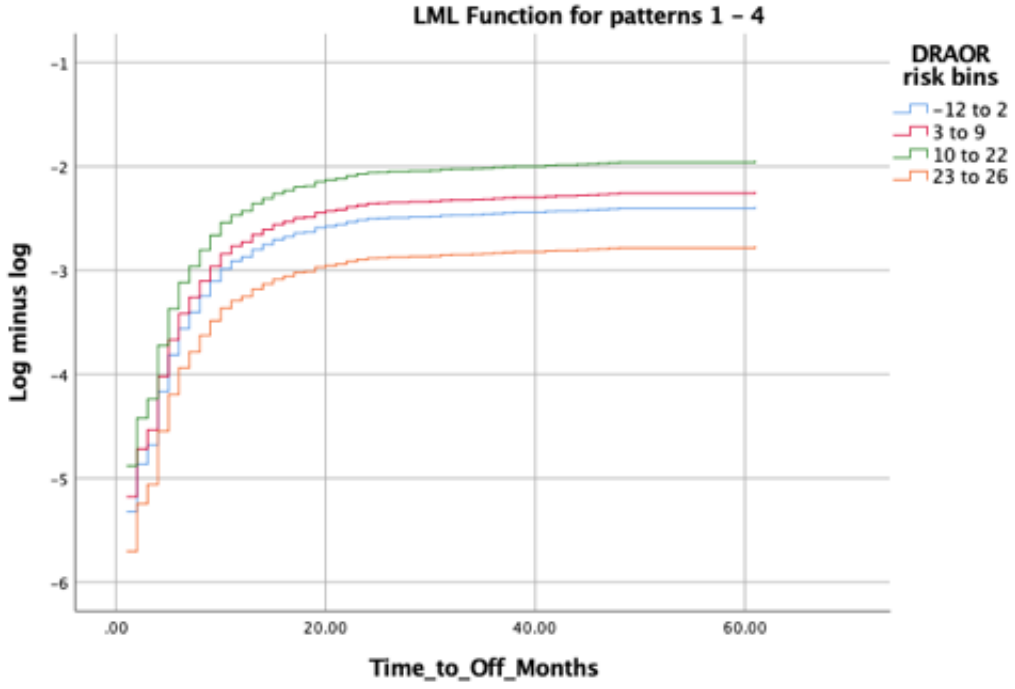


Figure E8. *New Offence by DRAOR Total Score.*

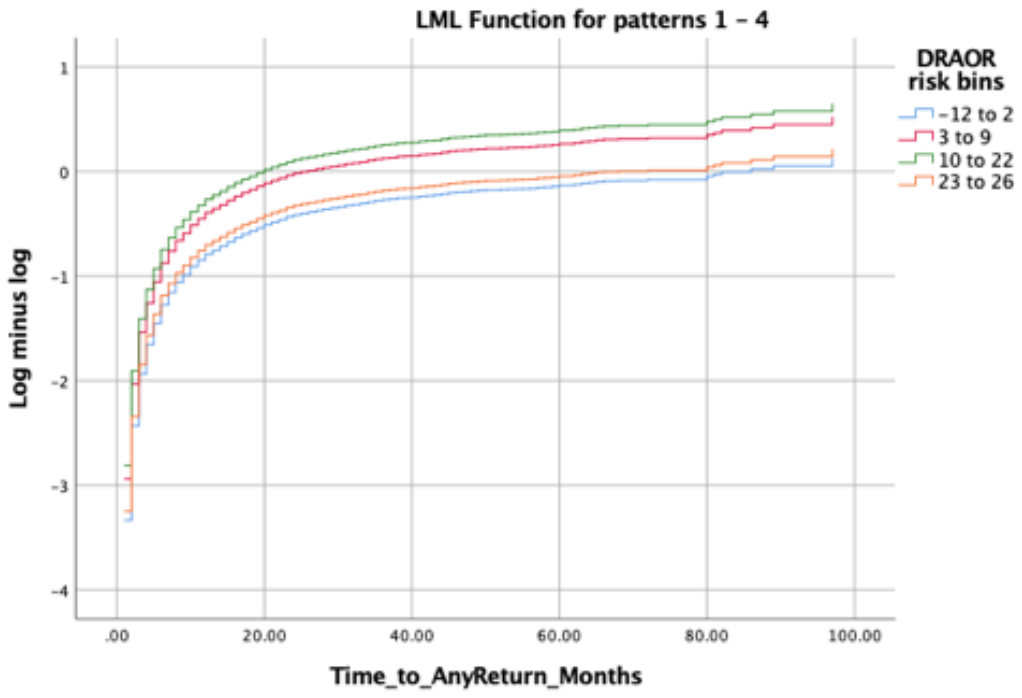


Figure E9. *Any Return by DRAOR Total Score.*

Appendix F: Tests of the Proportionality Assumption for Cox Regression

Table F1

Tests of Interactions Between Time and Levels of Each Covariate Predicting Technical Violations, New Offences, and Any Return

		B	SE	Wald	df	Sig.
	Race					
	White x time to violation [R]			.379	2	.827
	Black x time to violation	-.002	.011	.026	1	.873
	Hispanic x time to violation	.001	.011	.008	1	.930
	Level of Supervision					
	Administrative x time to violation [R]			2.447	4	.654
Technical	Minimum x time to violation	.009	.023	.141	1	.707
Violation	Low Normal x time to violation	.014	.013	1.182	1	.277
	High Normal x time to violation	-.002	.005	.121	1	.728
	Intensive x time to violation	.003	.005	.388	1	.533
	DRAOR Total Score Bins					
	Low–Moderate x time to violation [R]			5.107	3	.164
	Moderate x time to violation	-.003	.015	.031	1	.860
	Moderate–High x time to violation	-.013	.015	.766	1	.382
	High x time to violation	-.007	.015	.215	1	.643

Table F1 *continued.*

		B	SE	Wald	df	Sig.
	Race					
	White x time to offence [R]			.271	2	.873
	Black x time to offence	.000	.040	.000	1	.992
	Hispanic x time to offence	-.008	.042	.036	1	.850
	Level of Supervision					
	Administrative x time to offence [R]			.710	4	.950
	Minimum x time to offence	-.025	9.177	.000	1	.998
New Offence	Low Normal x time to offence	.020	.069	.088	1	.767
	High Normal x time to offence	-.013	.018	.502	1	.479
	Intensive x time to offence	.000	.014	.000	1	.984
	DRAOR Total Score Bins					
	Low–Moderate x time to offence [R]			7.825	3	.050
	Moderate x time to offence	.117	.130	.804	1	.370
	Moderate–High x time to offence	.088	.130	.460	1	.498
	High x time to offence	.071	.130	.300	1	.584

Table F1 *continued.*

		B	SE	Wald	df	Sig.
	Race					
	White x time to return [R]			.717	2	.699
	Black x time to return	-.004	.011	.126	1	.723
	Hispanic x time to return	.000	.012	.001	1	.979
	Level of Supervision					
	Administrative x time to return [R]			2.546	4	.636
	Minimum x time to return	.007	.023	.093	1	.760
Any Return	Low Normal x time to return	.013	.012	1.058	1	.304
	High Normal x time to return	-.004	.005	.466	1	.495
	Intensive x time to return	.001	.005	.052	1	.819
	DRAOR Total Score Bins					
	Low–Moderate x time to return [R]			5.419	3	.144
	Moderate x time to return	-.001	.016	.006	1	.937
	Moderate–High x time to return	-.012	.015	.633	1	.426
	High x time to return	-.006	.015	.137	1	.712

Note. [R] denotes the reference group; a *p* value of .006 was used to evaluate significance based on a Bonferroni correction to adjust for family wise error; none of the interaction terms achieved significance.

Appendix G: Recidivism Rates for White, Black and Hispanic Offenders by Supervision Level and Risk

Table G1

Base Rates for Each Outcome Disaggregated by Race and Level of Supervision

		Administrative		Minimum		Low Normal		High Normal		Intensive	
		n	Base rate	*n*	Base rate	*n*	Base rate	*n*	Base rate	*n*	Base rate
Technical Violation	White		†	24	46.2%	376	52.1%	487	69.3%	699	76.1%
	Black		†		†	106	**72.6%**	121	**77.1%**	181	**72.7%**
	Hispanic		†		†	15	48.4%	25	64.1%	29	70.7%
New Offence	White		†		†	43	6.0%	79	11.2%	142	15.5%
	Black		†		†	8	5.5%	11	7.0%	38	15.3%
	Hispanic		†		†		†		†	6	14.6%
Any Return	White		†	25	48.1%	378	52.4%	502	71.4%	722	78.6%
	Black		†		†	107	**73.3%**	121	**77.1%**	189	**75.9%**
	Hispanic		†		†	15	48.4%	25	64.1%	31	77.5%

Note. † = cells with *n* of 5 or less are supressed, bold denotes findings of particular interest.

Table G2

Base Rates for Each Outcome Disaggregated by Race and DRAOR Scores

		Low – Moderate		Moderate		Moderate – High		High	
		n	Base rate	*n*	Base rate	*n*	Base rate	*n*	Base rate
Technical Violation	White	324	53.5%	569	67.3%	684	72.6%	14	**56.0%**
	Black	86	**71.7%**	177	**73.8%**	143	**74.5%**	7	87.5%
	Hispanic	14	50.0%	28	59.6%	28	75.7%		†
New Offence	White	59	**9.7%**	79	**9.3%**	125	13.3%		†
	Black	6	5.0%	27	11.3%	24	12.5%		†
	Hispanic		†	5	10.6%		†		†
Any Return	White	33	55.5%	582	68.9%	703	74.6%	14	56.0%
	Black	87	**72.5%**	181	**75.4%**	147	**76.6%**	7	87.5%
	Hispanic	14	50.0%	30	63.8%	28	77.8%		†

Note. † = cells with *n* of 5 or less are supressed, bold denotes findings of particular interest.

Appendix H: Absolute Predictive Accuracy (Calibration) Results by Subscale Score

Table H1

DRAOR Stable Domain Score Absolute Predictive Accuracy for Technical Violations – White Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	39	0.38	0.17	15.00	6.79	0.45	0.27	0.75
1	80	0.41	0.34	33.00	27.52	**0.83**	**0.59**	**1.17**
2	155	0.54	0.52	83.00	80.45	**0.97**	**0.78**	**1.20**
3	213	0.53	0.51	113.00	108.84	**0.96**	**0.80**	**1.16**
4	232	0.59	0.56	138.00	130.85	**0.95**	**0.80**	**1.12**
5	248	0.63	0.61	155.00	150.04	**0.97**	**0.83**	**1.13**
6	307	0.66	0.68	203.00	207.53	**1.02**	**0.89**	**1.17**
7	289	0.67	0.78	195.00	224.84	**1.15**	**1.00**	**1.33**
8	257	0.68	0.69	175.00	177.84	**1.02**	**0.88**	**1.18**
9	164	0.65	0.71	107.00	115.78	**1.08**	**0.90**	**1.31**
10	124	0.72	1.00	89.00	124.00	1.39	1.13	1.71
11	61	0.64	1.00	39.00	61.00	1.56	1.14	2.14
12	25	0.52	1.00	13.00	25.00	1.92	1.12	3.31
Total	2194.00	0.62	0.56	1358.00	1222.06	0.90	0.85	0.95

Note. **Bold** denotes non-significant E/O index.

Table H2

DRAOR Acute Domain Score Absolute Predictive Accuracy for Technical Violations – White Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	14.00	0.43	0.08	6.00	1.16	0.19	0.09	0.43
1	48.00	0.52	0.23	25.00	11.09	0.44	0.30	0.66
2	89.00	0.43	0.39	38.00	34.62	**0.91**	**0.66**	**1.25**
3	150.00	0.59	0.50	88.00	75.00	**0.85**	**0.69**	**1.05**
4	238.00	0.55	0.56	130.00	132.09	**1.02**	**0.86**	**1.21**
5	237.00	0.58	0.59	137.00	139.36	**1.02**	**0.86**	**1.20**
6	276.00	0.62	0.68	170.00	186.30	**1.10**	**0.94**	**1.27**
7	290.00	0.61	0.64	176.00	184.44	**1.05**	**0.90**	**1.21**
8	273.00	0.68	0.77	186.00	208.85	**1.12**	**0.97**	**1.30**
9	209.00	0.67	0.73	141.00	153.20	**1.09**	**0.92**	**1.28**
10	146.00	0.63	0.80	92.00	116.80	1.27	1.03	1.56
11	91.00	0.77	0.90	70.00	81.90	**1.17**	**0.93**	**1.48**
12	85.00	0.73	1.00	62.00	85.00	1.37	1.07	1.76
13	37.00	0.81	1.00	30.00	37.00	**1.23**	**0.86**	**1.76**
14	11.00	0.64	1.00	7.00	11.00	**1.57**	**0.75**	**3.30**
Total	2194.00	0.62	0.56	1358.00	1222.06	0.90	0.85	0.95

Note. **Bold** denotes non-significant E/O index.

Table H3

DRAOR Protective Domain Score Absolute Predictive Accuracy for Technical Violations – White Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	73.0	0.73	1.00	53.00	73.00	1.38	1.05	1.80
1	109.0	0.66	0.93	72.00	101.70	1.41	1.12	1.78
2	162.0	0.69	0.88	111.00	141.75	1.28	1.06	1.54
3	168.0	0.73	0.90	123.00	151.20	1.23	1.03	1.47
4	230.0	0.61	0.68	140.00	156.40	**1.12**	**0.95**	**1.32**
5	261.0	0.64	0.59	168.00	154.77	**0.92**	**0.79**	**1.07**
6	451.0	0.61	0.54	273.00	244.44	**0.90**	**0.80**	**1.01**
7	231.0	0.60	0.51	139.00	118.50	**0.85**	**0.72**	**1.01**
8	163.0	0.61	0.48	100.00	78.40	0.78	0.64	0.95
9	121.0	0.57	0.44	69.00	53.00	0.77	0.61	0.97
10	83.00	0.55	0.39	46.00	32.45	0.71	0.53	0.94
11	73.00	0.45	0.21	33.00	15.18	0.46	0.33	0.65
12	69.00	0.45	0.12	31.00	8.35	0.27	0.19	0.38
Total	2194.0	0.62	0.56	1358.00	1222.06	0.90	0.85	0.95

Note. **Bold** denotes non-significant E/O index.

Table H4

DRAOR Total Score Absolute Predictive Accuracy for Technical Violations – White Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	558.00	0.48	0.37	270.00	205.34	0.76	0.68	0.86
Moderate	820.00	0.64	0.69	524.00	561.70	**1.07**	**0.98**	**1.17**
Mod.-High	802.00	0.69	0.82	553.00	660.85	1.20	1.10	1.30
High	14.00	0.79	0.95	11.00	13.30	**1.21**	**0.67**	**2.18**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H5

DRAOR Stable Domain Score Absolute Predictive Accuracy for New Offences – White Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	39.00	0.08	0.13	3.00	5.07	**1.69**	**0.55**	**5.24**
1	80.00	0.10	0.19	8.00	15.04	**1.88**	**0.94**	**3.76**
2	155.00	0.07	0.37	11.00	57.35	5.21	2.89	9.41
3	213.00	0.13	0.34	27.00	72.42	2.68	1.84	3.91
4	232.00	0.12	0.27	28.00	63.34	2.26	1.56	3.28
5	248.00	0.11	0.28	28.00	69.19	2.47	1.71	3.58
6	307.00	0.11	0.35	35.00	108.06	3.09	2.22	4.30
7	289.00	0.11	0.37	32.00	106.93	3.34	2.36	4.73
8	257.00	0.13	0.31	34.00	79.16	2.33	1.66	3.26
9	164.00	0.16	0.35	26.00	57.89	2.23	1.52	3.27
10	124.00	0.24	0.20	30.00	24.80	**0.83**	**0.58**	**1.18**
11	61.00	0.20	0.50	12.00	30.50	2.54	1.44	4.48
12	25.00	0.20	0.34	5.00	8.50	**1.70**	**0.71**	**4.08**
Total	2194.00	0.13	0.30	279.00	664.78	2.38	2.12	2.68

Note. **Bold** denotes non-significant E/O index.

Table H6

DRAOR Acute Domain Score Absolute Predictive Accuracy for New Offences – White Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	14.00	0.07	0.08	1.00	1.16	**1.16**	**0.16**	**8.25**
1	48.00	0.10	0.08	5.00	3.70	**0.74**	**0.31**	**1.78**
2	89.00	0.04	0.13	4.00	11.84	2.96	1.11	7.88
3	150.00	0.15	0.13	22.00	19.80	**0.90**	**0.59**	**1.37**
4	238.00	0.11	0.27	25.00	64.97	2.60	1.76	3.85
5	237.00	0.11	0.35	27.00	83.66	3.10	2.12	4.52
6	276.00	0.09	0.48	26.00	131.10	5.04	3.43	7.41
7	290.00	0.12	0.34	35.00	98.89	2.83	2.03	3.94
8	273.00	0.19	0.41	51.00	112.48	2.21	1.68	2.90
9	209.00	0.17	0.27	35.00	56.43	1.61	1.16	2.25
10	146.00	0.16	0.20	23.00	29.20	**1.27**	**0.84**	**1.91**
11	91.00	0.14	0.60	13.00	54.60	4.20	2.44	7.23
12	85.00	0.06	0.67	5.00	56.70	11.34	4.72	27.24
13	37.00	0.14	1.00	5.00	37.00	7.40	3.08	17.78
14	11.00	0.18	1.00	2.00	11.00	5.50	1.38	21.99
Total	2194.00	0.13	0.30	279.00	664.78	2.38	2.12	2.68

Note. **Bold** denotes non-significant E/O index.

Table H7

DRAOR Protective Domain Score Absolute Predictive Accuracy for New Offences – White Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	73.00	0.08	0.20	6.00	14.60	2.43	1.09	5.42
1	109.00	0.17	0.67	18.00	72.70	4.04	2.54	6.41
2	162.00	0.15	0.50	25.00	81.00	3.24	2.19	4.79
3	168.00	0.17	0.40	29.00	67.20	2.32	1.61	3.33
4	230.00	0.14	0.32	32.00	73.60	2.30	1.63	3.25
5	261.00	0.11	0.30	28.00	77.26	2.76	1.91	4.00
6	451.00	0.13	0.36	60.00	160.11	2.67	2.07	3.44
7	231.00	0.10	0.23	24.00	53.36	2.22	1.49	3.32
8	163.00	0.13	0.30	22.00	48.25	2.19	1.44	3.33
9	121.00	0.14	0.31	17.00	37.87	2.23	1.38	3.58
10	83.00	0.06	0.48	5.00	39.67	7.93	3.30	19.06
11	73.00	0.11	0.21	8.00	15.18	**1.90**	**0.95**	**3.80**
12	69.00	0.07	0.06	5.00	4.21	**0.84**	**0.35**	**2.02**
Total	2194.0	0.13	0.30	279.00	664.78	2.38	2.12	2.68

Note. **Bold** denotes non-significant E/O index.

Table H8

DRAOR Total Score Absolute Predictive Accuracy for New Offences – White Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	558.00	0.09	0.23	52.00	130.01	2.50	1.91	3.28
Moderate	820.00	0.12	0.38	100.00	309.96	3.10	2.55	3.77
Mod.-High	802.00	0.16	0.33	125.00	267.07	2.14	1.79	2.55
High	14.00	0.14	0.50	2.00	7.00	**3.50**	**0.88**	**13.99**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H9

DRAOR Stable Domain Score Absolute Predictive Accuracy for Any Return – White Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	39.00	0.38	0.22	15.00	8.46	0.56	0.34	0.94
1	80.00	0.43	0.34	34.00	27.52	**0.81**	**0.58**	**1.13**
2	155.00	0.54	0.59	84.00	91.92	**1.09**	**0.88**	**1.36**
3	213.00	0.54	0.62	115.00	131.42	**1.14**	**0.95**	**1.37**
4	232.00	0.61	0.49	141.00	112.52	0.80	0.68	0.94
5	248.00	0.65	0.65	161.00	161.45	**1.00**	**0.86**	**1.17**
6	307.00	0.69	0.75	211.00	229.02	**1.09**	**0.95**	**1.24**
7	289.00	0.69	0.85	198.00	246.23	1.24	1.08	1.43
8	257.00	0.72	0.69	184.00	177.84	**0.97**	**0.84**	**1.12**
9	164.00	0.69	0.77	113.00	125.46	**1.11**	**0.92**	**1.34**
10	124.00	0.77	1.00	95.00	124.00	1.31	1.07	1.60
11	61.00	0.70	1.00	43.00	61.00	1.42	1.05	1.91
12	25.00	0.56	1.00	14.00	25.00	1.79	1.06	3.02
Total	2194.00	0.64	0.62	1408.00	1362.47	**0.97**	**0.92**	**1.02**

Note. **Bold** denotes non-significant E/O index.

Table H10

DRAOR Acute Domain Score Absolute Predictive Accuracy for Any Return – White Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	14.00	0.43	0.17	6.00	2.34	0.39	0.18	0.87
1	48.00	0.52	0.27	25.00	12.91	0.52	0.35	0.76
2	89.00	0.45	0.44	40.00	39.52	**0.99**	**0.72**	**1.35**
3	150.00	0.60	0.50	90.00	75.00	**0.83**	**0.68**	**1.02**
4	238.00	0.55	0.50	131.00	119.00	**0.91**	**0.77**	**1.08**
5	237.00	0.59	0.68	140.00	160.21	**1.14**	**0.97**	**1.35**
6	276.00	0.63	0.80	175.00	220.80	1.26	1.09	1.46
7	290.00	0.63	0.75	184.00	217.50	1.18	1.02	1.37
8	273.00	0.72	0.85	196.00	232.87	1.19	1.03	1.37
9	209.00	0.72	0.73	151.00	153.20	**1.01**	**0.86**	**1.19**
10	146.00	0.68	1.00	99.00	146.00	1.47	1.21	1.80
11	91.00	0.78	0.80	71.00	72.80	**1.03**	**0.81**	**1.29**
12	85.00	0.73	1.00	62.00	85.00	1.37	1.07	1.76
13	37.00	0.81	1.00	30.00	37.00	**1.23**	**0.86**	**1.76**
14	11.00	0.73	1.00	8.00	11.00	**1.38**	**0.69**	**2.75**
Total	2194.00	0.64	0.62	1408.00	1362.47	**0.97**	**0.92**	**1.02**

Note. **Bold** denotes non-significant E/O index.

Table H11

DRAOR Protective Domain Score Absolute Predictive Accuracy for Any Return – White Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	73.00	0.74	1.00	54.00	73.00	1.35	1.04	1.77
1	109.00	0.69	0.67	75.00	72.70	**0.97**	**0.77**	**1.22**
2	162.00	0.73	0.88	118.00	141.75	1.20	1.00	1.44
3	168.00	0.77	0.90	129.00	151.20	**1.17**	**0.99**	**1.39**
4	230.00	0.66	0.72	151.00	165.60	**1.10**	**0.94**	**1.29**
5	261.00	0.65	0.67	169.00	174.09	**1.03**	**0.89**	**1.20**
6	451.00	0.63	0.80	284.00	359.00	1.26	1.13	1.42
7	231.00	0.61	0.59	142.00	136.29	**0.96**	**0.81**	**1.13**
8	163.00	0.63	0.52	102.00	84.60	**0.83**	**0.68**	**1.01**
9	121.00	0.59	0.50	71.00	60.50	**0.85**	**0.68**	**1.08**
10	83.00	0.55	0.61	46.00	50.55	**1.10**	**0.82**	**1.47**
11	73.00	0.49	0.29	36.00	21.32	0.59	0.43	0.82
12	69.00	0.45	0.15	31.00	10.49	0.34	0.24	0.48
Total	2194.00	0.64	0.62	1408.00	1362.47	**0.97**	**0.92**	**1.02**

Note. **Bold** denotes non-significant E/O index.

Table H12

DRAOR Total Score Absolute Predictive Accuracy for Any Return – White Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	558.00	0.50	0.43	277.00	239.38	0.86	0.77	0.97
Moderate	820.00	0.65	0.76	535.00	626.48	1.17	1.08	1.27
Mod.-High	802.00	0.73	0.86	585.00	692.13	1.18	1.09	1.28
High	14.00	0.79	1.00	11.00	14.00	**1.27**	**0.70**	**2.30**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H13

DRAOR Stable Domain Score Absolute Predictive Accuracy for Technical Violations – Black Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	5.00	0.20	0.17	1.00	0.87	**0.87**	**0.12**	**6.18**
1	3.00	0.33	0.34	1.00	1.03	**1.03**	**0.15**	**7.33**
2	23.00	0.65	0.52	15.00	11.94	**0.80**	**0.48**	**1.32**
3	45.00	0.58	0.51	26.00	23.00	**0.88**	**0.60**	**1.30**
4	55.00	0.53	0.56	29.00	31.02	**1.07**	**0.74**	**1.54**
5	69.00	0.62	0.61	43.00	41.75	**0.97**	**0.72**	**1.31**
6	70.00	0.71	0.68	50.00	47.32	**0.95**	**0.72**	**1.25**
7	61.00	0.66	0.78	40.00	47.46	**1.19**	**0.87**	**1.62**
8	49.00	0.63	0.69	31.00	33.91	**1.09**	**0.77**	**1.56**
9	51.00	0.55	0.71	28.00	36.01	**1.29**	**0.89**	**1.86**
10	36.00	0.72	1.00	26.00	36.00	**1.38**	**0.94**	**2.03**
11	14.00	0.50	1.00	7.00	14.00	**2.00**	**0.95**	**4.20**
12	10.00	0.50	1.00	5.00	10.00	**2.00**	**0.83**	**4.81**
Total	491.00	0.62	0.56	302.00	273.49	**0.91**	**0.81**	**1.01**

Note. **Bold** denotes non-significant E/O index.

Table H14

DRAOR Acute Domain Score Absolute Predictive Accuracy for Technical Violations – Black Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	1.00	0.08	1.00	0.08	0.08	0.01	0.59
1	7.00	0.57	0.23	4.00	1.62	**0.40**	**0.15**	**1.08**
2	16.00	0.56	0.39	9.00	6.22	**0.69**	**0.36**	**1.33**
3	32.00	0.53	0.50	17.00	16.00	**0.94**	**0.59**	**1.51**
4	45.00	0.67	0.56	30.00	24.98	**0.83**	**0.58**	**1.19**
5	64.00	0.66	0.59	42.00	37.63	**0.90**	**0.66**	**1.21**
6	70.00	0.69	0.68	48.00	47.25	**0.98**	**0.74**	**1.31**
7	71.00	0.61	0.64	43.00	45.16	**1.05**	**0.78**	**1.42**
8	78.00	0.55	0.77	43.00	59.67	1.39	1.03	1.87
9	36.00	0.53	0.73	19.00	26.39	**1.39**	**0.89**	**2.18**
10	30.00	0.73	0.80	22.00	24.00	**1.09**	**0.72**	**1.66**
11	25.00	0.48	0.90	12.00	22.50	1.88	1.06	3.30
12	13.00	0.69	1.00	9.00	13.00	**1.44**	**0.75**	**2.78**
13	2.00	1.00	1.00	2.00	2.00	**1.00**	**0.25**	**4.00**
14	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
Total	491.00	0.62	0.56	302.00	273.49	**0.91**	**0.81**	**1.01**

Note. **Bold** denotes non-significant E/O index.

Table H15

DRAOR Protective Domain Score Absolute Predictive Accuracy for Technical Violations – Black Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	20.00	0.50	1.00	10.00	20.00	2.00	1.08	3.72
1	22.00	0.82	0.93	18.00	20.53	**1.14**	**0.72**	**1.81**
2	33.00	0.70	0.88	23.00	28.88	**1.26**	**0.83**	**1.89**
3	43.00	0.65	0.90	28.00	38.70	**1.38**	**0.95**	**2.00**
4	50.00	0.64	0.68	32.00	34.00	**1.06**	**0.75**	**1.50**
5	52.00	0.54	0.59	28.00	30.84	**1.10**	**0.76**	**1.60**
6	119.00	0.59	0.54	70.00	64.50	**0.92**	**0.73**	**1.16**
7	48.00	0.52	0.51	25.00	24.62	**0.98**	**0.67**	**1.46**
8	40.00	0.68	0.48	27.00	19.24	**0.71**	**0.49**	**1.04**
9	25.00	0.64	0.44	16.00	10.95	**0.68**	**0.42**	**1.12**
10	22.00	0.64	0.39	14.00	8.60	**0.61**	**0.36**	**1.04**
11	7.00	0.43	0.21	3.00	1.46	**0.49**	**0.16**	**1.50**
12	10.00	0.80	0.12	8.00	1.21	0.15	0.08	0.30
Total	491.00	0.62	0.56	302.00	273.49	**0.91**	**0.81**	**1.01**

Note. **Bold** denotes non-significant E/O index.

Table H16

DRAOR Total Score Absolute Predictive Accuracy for Technical Violations – Black Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	98.00	0.58	0.37	57.00	36.06	0.63	0.49	0.82
Moderate	223.00	0.64	0.69	142.00	152.76	**1.08**	**0.91**	**1.27**
Mod.-High	166.00	0.60	0.82	99.00	136.78	1.38	1.13	1.68
High	4.00	1.00	0.95	4.00	3.80	**0.95**	**0.36**	**2.53**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H17

DRAOR Stable Domain Score Absolute Predictive Accuracy for New Offence – Black Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	5.00	0.00	0.13	0.00	0.65	UtC	UtC	UtC
1	3.00	0.33	0.19	1.00	0.56	**0.56**	**0.08**	**4.00**
2	23.00	0.30	0.37	7.00	8.51	**1.22**	**0.58**	**2.55**
3	45.00	0.11	0.34	5.00	15.30	3.06	1.27	7.35
4	55.00	0.18	0.27	10.00	15.02	**1.50**	**0.81**	**2.79**
5	69.00	0.09	0.28	6.00	19.25	3.21	1.44	7.14
6	70.00	0.20	0.35	14.00	24.64	1.76	1.04	2.97
7	61.00	0.16	0.37	10.00	22.57	2.26	1.21	4.19
8	49.00	0.18	0.31	9.00	15.09	**1.68**	**0.87**	**3.22**
9	51.00	0.16	0.35	8.00	18.00	2.25	1.13	4.50
10	36.00	0.17	0.20	6.00	7.20	**1.20**	**0.54**	**2.67**
11	14.00	0.36	0.50	5.00	7.00	**1.40**	**0.58**	**3.36**
12	10.00	0.50	0.34	5.00	3.40	**0.68**	**0.28**	**1.63**
Total	491.00	0.18	0.30	86.00	148.77	**1.73**	**1.40**	**2.14**

Note. **Bold** denotes non-significant E/O index.

Table H18

DRAOR Acute Domain Score Absolute Predictive Accuracy for New Offences – Black Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.08	0.00	0.08	UtC	UtC	UtC
1	7.00	0.29	0.54	2.00	0.54	**0.27**	**0.07**	**1.08**
2	16.00	0.19	2.13	3.00	2.13	**0.71**	**0.23**	**2.20**
3	32.00	0.13	4.22	4.00	4.22	**1.06**	**0.40**	**2.81**
4	45.00	0.09	12.29	4.00	12.29	3.07	1.15	8.18
5	64.00	0.17	22.59	11.00	22.59	2.05	1.14	3.71
6	70.00	0.20	33.25	14.00	33.25	2.38	1.41	4.01
7	71.00	0.14	24.21	10.00	24.21	2.42	1.30	4.50
8	78.00	0.24	32.14	19.00	32.14	1.69	1.08	2.65
9	36.00	0.25	9.72	9.00	9.72	**1.08**	**0.56**	**2.08**
10	30.00	0.07	6.00	2.00	6.00	**3.00**	**0.75**	**12.00**
11	25.00	0.16	15.00	4.00	15.00	3.75	1.41	9.99
12	13.00	0.31	8.67	4.00	8.67	**2.17**	**0.81**	**5.78**
13	2.00	0.00	2.00	0.00	2.00	UtC	UtC	UtC
14	1.00	0.00	1.00	0.00	1.00	UtC	UtC	UtC
Total	491.00	0.18	148.77	86.00	148.77	1.73	1.40	2.14

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H19

DRAOR Protective Domain Score Absolute Predictive Accuracy for New Offences – Black Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	20.00	0.15	0.20	3.00	4.00	**1.33**	**0.43**	**4.13**
1	22.00	0.32	0.67	7.00	14.67	**2.10**	**1.00**	**4.40**
2	33.00	0.12	0.50	4.00	16.50	4.13	1.55	10.99
3	43.00	0.16	0.40	7.00	17.20	2.46	1.17	5.15
4	50.00	0.26	0.32	13.00	16.00	**1.23**	**0.71**	**2.12**
5	52.00	0.19	0.30	10.00	15.39	**1.54**	**0.83**	**2.86**
6	119.00	0.16	0.36	19.00	42.25	2.22	1.42	3.49
7	48.00	0.21	0.23	10.00	11.09	**1.11**	**0.60**	**2.06**
8	40.00	0.15	0.30	6.00	11.84	**1.97**	**0.89**	**4.39**
9	25.00	0.04	0.31	1.00	7.83	7.83	1.10	55.55
10	22.00	0.18	0.48	4.00	10.52	**2.63**	**0.99**	**7.00**
11	7.00	0.14	0.21	1.00	1.46	**1.46**	**0.21**	**10.34**
12	10.00	0.10	0.06	1.00	0.61	**0.61**	**0.09**	**4.33**
Total	491.00	0.18	0.30	86.00	148.77	1.73	1.40	2.14

Note. **Bold** denotes non-significant E/O index.

Table H20

DRAOR Total Score Absolute Predictive Accuracy for New Offence –Black Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	98.00	0.16	0.23	16.00	22.83	**1.43**	**0.87**	**2.33**
Moderate	223.00	0.15	0.38	34.00	84.29	2.48	1.77	3.47
Mod.-High	166.00	0.20	0.33	33.00	55.28	1.68	1.19	2.36
High	4.00	0.75	0.50	3.00	2.00	**0.67**	**0.22**	**2.07**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H21

DRAOR Stable Domain Score Absolute Predictive Accuracy for Any Return – Black Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	5.00	0.20	0.22	1.00	1.09	**1.09**	**0.15**	**7.70**
1	3.00	0.33	0.34	1.00	1.03	**1.03**	**0.15**	**7.33**
2	23.00	0.70	0.59	16.00	13.64	**0.85**	**0.52**	**1.39**
3	45.00	0.60	0.62	27.00	27.77	**1.03**	**0.71**	**1.50**
4	55.00	0.55	0.49	30.00	26.68	**0.89**	**0.62**	**1.27**
5	69.00	0.64	0.65	44.00	44.92	**1.02**	**0.76**	**1.37**
6	70.00	0.76	0.75	53.00	52.22	**0.99**	**0.75**	**1.29**
7	61.00	0.72	0.85	44.00	51.97	**1.18**	**0.88**	**1.59**
8	49.00	0.73	0.69	36.00	33.91	**0.94**	**0.68**	**1.31**
9	51.00	0.59	0.77	30.00	39.02	**1.30**	**0.91**	**1.86**
10	36.00	0.72	1.00	26.00	36.00	**1.38**	**0.94**	**2.03**
11	14.00	0.57	1.00	8.00	14.00	**1.75**	**0.88**	**3.50**
12	10.00	0.60	1.00	6.00	10.00	**1.67**	**0.75**	**3.71**
Total	491.00	0.66	0.62	322.00	304.91	**0.95**	**0.85**	**1.06**

Note. **Bold** denotes non-significant E/O index.

Table H22

DRAOR Acute Domain Score Absolute Predictive Accuracy for Any Return – Black Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	1.00	0.17	1.00	0.17	**0.17**	**0.02**	**1.19**
1	7.00	0.57	0.27	4.00	1.88	**0.47**	**0.18**	**1.25**
2	16.00	0.56	0.44	9.00	7.10	**0.79**	**0.41**	**1.52**
3	32.00	0.53	0.50	17.00	16.00	**0.94**	**0.59**	**1.51**
4	45.00	0.67	0.50	30.00	22.50	**0.75**	**0.52**	**1.07**
5	64.00	0.67	0.68	43.00	43.26	**1.01**	**0.75**	**1.36**
6	70.00	0.71	0.80	50.00	56.00	**1.12**	**0.85**	**1.48**
7	71.00	0.65	0.75	46.00	53.25	**1.16**	**0.87**	**1.55**
8	78.00	0.67	0.85	52.00	66.53	**1.28**	**0.97**	**1.68**
9	36.00	0.61	0.73	22.00	26.39	**1.20**	**0.79**	**1.82**
10	30.00	0.73	1.00	22.00	30.00	**1.36**	**0.90**	**2.07**
11	25.00	0.56	0.80	14.00	20.00	**1.43**	**0.85**	**2.41**
12	13.00	0.69	1.00	9.00	13.00	**1.44**	**0.75**	**2.78**
13	2.00	1.00	1.00	2.00	2.00	**1.00**	**0.25**	**4.00**
14	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
Total	491.00	0.66	0.62	322.00	304.91	**0.95**	**0.85**	**1.06**

Note. **Bold** denotes non-significant E/O index.

Table H23

DRAOR Protective Domain Score Absolute Predictive Accuracy for Any Return – Black Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	20.00	0.50	1.00	10.00	20.00	2.00	1.08	3.72
1	22.00	0.82	0.67	18.00	14.67	**0.82**	**0.51**	**1.29**
2	33.00	0.70	0.88	23.00	28.88	**1.26**	**0.83**	**1.89**
3	43.00	0.67	0.90	29.00	38.70	**1.33**	**0.93**	**1.92**
4	50.00	0.72	0.72	36.00	36.00	**1.00**	**0.72**	**1.39**
5	52.00	0.62	0.67	32.00	34.68	**1.08**	**0.77**	**1.53**
6	119.00	0.65	0.80	77.00	94.72	**1.23**	**0.98**	**1.54**
7	48.00	0.58	0.59	28.00	28.32	**1.01**	**0.70**	**1.46**
8	40.00	0.70	0.52	28.00	20.76	**0.74**	**0.51**	**1.07**
9	25.00	0.64	0.50	16.00	12.50	**0.78**	**0.48**	**1.28**
10	22.00	0.64	0.61	14.00	13.40	**0.96**	**0.57**	**1.62**
11	7.00	0.43	0.29	3.00	2.04	**0.68**	**0.22**	**2.11**
12	10.00	0.80	0.15	8.00	1.52	0.19	0.10	0.38
Total	491.00	0.66	0.62	322.00	304.91	**0.95**	**0.85**	**1.06**

Note. **Bold** denotes non-significant E/O index.

Table H24

DRAOR Total Score Absolute Predictive Accuracy for Any Return – Black Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	98.00	0.59	0.43	58.00	42.04	0.72	0.56	0.94
Moderate	223.00	0.68	0.76	151.00	170.37	**1.13**	**0.96**	**1.32**
Mod.-High	166.00	0.66	0.86	109.00	143.26	1.31	1.09	1.59
High	4.00	1.00	1.00	4.00	4.00	**1.00**	**0.38**	**2.66**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H25

DRAOR Stable Domain Score Absolute Predictive Accuracy for Technical Violations – Hispanic Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.17	0.00	0.17	UtC	UtC	UtC
1	3.00	0.33	0.34	1.00	1.03	**1.03**	**0.15**	**7.33**
2	6.00	0.50	0.52	3.00	3.11	**1.04**	**0.33**	**3.22**
3	7.00	0.57	0.51	4.00	3.58	**0.89**	**0.34**	**2.38**
4	10.00	0.40	0.56	4.00	5.64	**1.41**	**0.53**	**3.76**
5	10.00	0.60	0.61	6.00	6.05	**1.01**	**0.45**	**2.24**
6	15.00	0.73	0.68	11.00	10.14	**0.92**	**0.51**	**1.66**
7	8.00	1.00	0.78	8.00	6.22	**0.78**	**0.39**	**1.56**
8	8.00	0.75	0.69	6.00	5.54	**0.92**	**0.41**	**2.05**
9	4.00	1.00	0.71	4.00	2.82	**0.71**	**0.26**	**1.88**
10	2.00	0.50	1.00	1.00	2.00	**2.00**	**0.28**	**14.20**
11	3.00	1.00	1.00	3.00	3.00	**1.00**	**0.32**	**3.10**
12	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
Total	78.00	0.67	0.56	52.00	43.45	**0.84**	**0.64**	**1.10**

Note. **Bold** denotes non-significant E/O index.

Table H26

DRAOR Acute Domain Score Absolute Predictive Accuracy for Technical Violations – Hispanic Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.08	0.00	0.08	UtC	UtC	UtC
1	6.00	0.50	0.23	3.00	1.39	**0.46**	**0.15**	**1.43**
2	4.00	0.50	0.39	2.00	1.56	**0.78**	**0.19**	**3.11**
3	6.00	0.67	0.50	4.00	3.00	**0.75**	**0.28**	**2.00**
4	10.00	0.80	0.56	8.00	5.55	**0.69**	**0.35**	**1.39**
5	6.00	0.67	0.59	4.00	3.53	**0.88**	**0.33**	**2.35**
6	5.00	0.40	0.68	2.00	3.38	**1.69**	**0.42**	**6.75**
7	12.00	0.75	0.64	9.00	7.63	**0.85**	**0.44**	**1.63**
8	11.00	0.73	0.77	8.00	8.42	**1.05**	**0.53**	**2.10**
9	10.00	0.80	0.73	8.00	7.33	**0.92**	**0.46**	**1.83**
10	3.00	1.00	0.80	3.00	2.40	**0.80**	**0.26**	**2.48**
11	2.00	0.00	0.90	0.00	1.80	UtC	UtC	UtC
12	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
13	1.00	0.00	1.00	0.00	1.00	UtC	UtC	UtC
14	0.00	UtC	1.00	0.00	0.00	UtC	UtC	UtC
Total	78.00	0.67	0.56	52.00	43.45	**0.84**	**0.64**	**1.10**

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H27

DRAOR Protective Domain Score Absolute Predictive Accuracy for Technical Violations – Hispanic Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
1	2.00	1.00	0.93	2.00	1.87	**0.93**	**0.23**	**3.73**
2	3.00	0.67	0.88	2.00	2.63	**1.31**	**0.33**	**5.25**
3	7.00	0.71	0.90	5.00	6.30	**1.26**	**0.52**	**3.03**
4	9.00	0.89	0.68	8.00	6.12	**0.77**	**0.38**	**1.53**
5	6.00	0.67	0.59	4.00	3.56	**0.89**	**0.33**	**2.37**
6	27.00	0.63	0.54	17.00	14.63	**0.86**	**0.54**	**1.38**
7	9.00	0.56	0.51	5.00	4.62	**0.92**	**0.38**	**2.22**
8	7.00	0.43	0.48	3.00	3.37	**1.12**	**0.36**	**3.48**
9	1.00	1.00	0.44	1.00	0.44	**0.44**	**0.06**	**3.11**
10	4.00	0.50	0.39	2.00	1.56	**0.78**	**0.20**	**3.13**
11	2.00	1.00	0.21	2.00	0.42	0.21	0.05	0.83
12	0.00	UtC	0.12	0.00	0.00	UtC	UtC	UtC
Total	78.00	0.67	0.56	52.00	43.45	**0.84**	**0.64**	**1.10**

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H28

DRAOR Total Score Absolute Predictive Accuracy for Technical Violation – Hispanic Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	22	0.45	0.37	10.00	8.10	**0.81**	**0.44**	**1.50**
Moderate	35	0.77	0.69	27.00	23.98	**0.89**	**0.61**	**1.29**
Mod.-High	21	0.71	0.82	15.00	17.30	**1.15**	**0.70**	**1.91**
High	0	UtC	0.95	0.00	0.00	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H29

DRAOR Stable Domain Score Absolute Predictive Accuracy for New Offences – Hispanic Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.13	0.00	0.13	UtC	UtC	UtC
1	3.00	0.00	0.19	0.00	0.56	UtC	UtC	UtC
2	6.00	0.17	0.37	1.00	2.22	**2.22**	**0.31**	**15.76**
3	7.00	0.14	0.34	1.00	2.38	**2.38**	**0.34**	**16.90**
4	10.00	0.30	0.27	3.00	2.73	**0.91**	**0.29**	**2.82**
5	10.00	0.00	0.28	0.00	2.79	UtC	UtC	UtC
6	15.00	0.27	0.35	4.00	5.28	**1.32**	**0.50**	**3.52**
7	8.00	0.13	0.37	1.00	2.96	**2.96**	**0.42**	**21.01**
8	8.00	0.00	0.31	0.00	2.46	UtC	UtC	UtC
9	4.00	0.25	0.35	1.00	1.41	**1.41**	**0.20**	**10.02**
10	2.00	0.00	0.20	0.00	0.40	UtC	UtC	UtC
11	3.00	0.00	0.50	0.00	1.50	UtC	UtC	UtC
12	1.00	0.00	0.34	0.00	0.34	UtC	UtC	UtC
Total	78.00	0.14	0.30	11.00	23.63	2.15	1.19	3.88

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H30

DRAOR Acute Domain Score Absolute Predictive Accuracy for New Offences – Hispanic Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.08	0.00	0.08	UtC	UtC	UtC
1	6.00	0.00	0.08	0.00	0.46	UtC	UtC	UtC
2	4.00	0.00	0.13	0.00	0.53	UtC	UtC	UtC
3	6.00	0.50	0.13	3.00	0.79	0.26	0.09	0.82
4	10.00	0.20	0.27	2.00	2.73	**1.37**	**0.34**	**5.46**
5	6.00	0.00	0.35	0.00	2.12	UtC	UtC	UtC
6	5.00	0.20	0.48	1.00	2.38	**2.38**	**0.33**	**16.86**
7	12.00	0.08	0.34	1.00	4.09	**4.09**	**0.58**	**29.05**
8	11.00	0.00	0.41	0.00	4.53	UtC	UtC	UtC
9	10.00	0.30	0.27	3.00	2.70	**0.90**	**0.29**	**2.79**
10	3.00	0.33	0.20	1.00	0.60	**0.60**	**0.08**	**4.26**
11	2.00	0.00	0.60	0.00	1.20	UtC	UtC	UtC
12	1.00	0.00	0.67	0.00	0.67	UtC	UtC	UtC
13	1.00	0.00	1.00	0.00	1.00	UtC	UtC	UtC
14	0.00	UtC	1.00	0.00	0.00	UtC	UtC	UtC
Total	78.00	0.14	0.30	11.00	23.63	2.15	1.19	3.88

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H31

DRAOR Protective Domain Score Absolute Predictive Accuracy for New Offences – Hispanic Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.20	0.00	0.20	UtC	UtC	UtC
1	2.00	0.50	0.67	1.00	1.33	**1.33**	**0.19**	**9.47**
2	3.00	0.00	0.50	0.00	1.50	UtC	UtC	UtC
3	7.00	0.14	0.40	1.00	2.80	**2.80**	**0.39**	**19.88**
4	9.00	0.22	0.32	2.00	2.88	**1.44**	**0.36**	**5.76**
5	6.00	0.17	0.30	1.00	1.78	**1.78**	**0.25**	**12.61**
6	27.00	0.15	0.36	4.00	9.59	**2.40**	**0.90**	**6.38**
7	9.00	0.22	0.23	2.00	2.08	**1.04**	**0.26**	**4.16**
8	7.00	0.00	0.30	0.00	2.07	UtC	UtC	UtC
9	1.00	0.00	0.31	0.00	0.31	UtC	UtC	UtC
10	4.00	0.00	0.48	0.00	1.91	UtC	UtC	UtC
11	0.00	UtC	0.21	0.00	0.00	UtC	UtC	UtC
12	2.00	0.00	0.06	0.00	0.12	UtC	UtC	UtC
Total	78.00	0.14	0.30	11.00	23.63	2.15	1.19	3.88

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H32

DRAOR Total Score Absolute Predictive Accuracy for New Offences – Hispanic Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	22	0.09	0.23	2.00	5.13	**2.56**	**0.64**	**10.25**
Moderate	35	0.17	0.38	6.00	13.23	**2.21**	**0.99**	**4.91**
Mod.-High	21	0.14	0.33	3.00	6.99	**2.33**	**0.75**	**7.23**
High	0	UtC	0.50	0.00	0.00	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H33

DRAOR Stable Domain Score Absolute Predictive Accuracy for Any Return – Hispanic Men

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.22	0.00	0.22	UtC	UtC	UtC
1	3.00	0.33	0.34	1.00	1.03	**1.03**	**0.15**	**7.33**
2	6.00	0.50	0.59	3.00	3.56	**1.19**	**0.38**	**3.68**
3	7.00	0.57	0.62	4.00	4.32	**1.08**	**0.41**	**2.88**
4	10.00	0.50	0.49	5.00	4.85	**0.97**	**0.40**	**2.33**
5	10.00	0.60	0.65	6.00	6.51	**1.09**	**0.49**	**2.42**
6	15.00	0.73	0.75	11.00	11.19	**1.02**	**0.56**	**1.84**
7	8.00	1.00	0.85	8.00	6.82	**0.85**	**0.43**	**1.70**
8	8.00	0.75	0.69	6.00	5.54	**0.92**	**0.41**	**2.05**
9	4.00	1.00	0.77	4.00	3.06	**0.77**	**0.29**	**2.04**
10	2.00	0.50	1.00	1.00	2.00	**2.00**	**0.28**	**14.20**
11	3.00	1.00	1.00	3.00	3.00	**1.00**	**0.32**	**3.10**
12	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
Total	78.00	0.68	0.62	53.00	48.44	**0.91**	**0.70**	**1.20**

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H34

DRAOR Acute Domain Score Absolute Predictive Accuracy for Any Return – Hispanic Men

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	0.00	0.17	0.00	0.17	UtC	UtC	UtC
1	6.00	0.50	0.27	3.00	1.61	**0.54**	**0.17**	**1.67**
2	4.00	0.50	0.44	2.00	1.78	**0.89**	**0.22**	**3.55**
3	6.00	0.67	0.50	4.00	3.00	**0.75**	**0.28**	**2.00**
4	10.00	0.80	0.50	8.00	5.00	**0.63**	**0.31**	**1.25**
5	6.00	0.67	0.68	4.00	4.06	**1.01**	**0.38**	**2.70**
6	5.00	0.40	0.80	2.00	4.00	**2.00**	**0.50**	**8.00**
7	12.00	0.75	0.75	9.00	9.00	**1.00**	**0.52**	**1.92**
8	11.00	0.73	0.85	8.00	9.38	**1.17**	**0.59**	**2.35**
9	10.00	0.90	0.73	9.00	7.33	**0.81**	**0.42**	**1.57**
10	3.00	1.00	1.00	3.00	3.00	**1.00**	**0.32**	**3.10**
11	2.00	0.00	0.80	0.00	1.60	UtC	UtC	UtC
12	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
13	1.00	0.00	1.00	0.00	1.00	UtC	UtC	UtC
14	0.00	UtC	1.00	0.00	0.00	UtC	UtC	UtC
Total	78.00	0.68	0.62	53.00	48.44	**0.91**	**0.70**	**1.20**

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H35

DRAOR Protective Domain Score Absolute Predictive Accuracy for Any Return – Hispanic Men

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
1	2.00	1.00	0.67	2.00	1.33	**0.67**	**0.17**	**2.67**
2	3.00	0.67	0.88	2.00	2.63	**1.31**	**0.33**	**5.25**
3	7.00	0.86	0.90	6.00	6.30	**1.05**	**0.47**	**2.34**
4	9.00	0.89	0.72	8.00	6.48	**0.81**	**0.41**	**1.62**
5	6.00	0.67	0.67	4.00	4.00	**1.00**	**0.38**	**2.67**
6	27.00	0.63	0.80	17.00	21.49	**1.26**	**0.79**	**2.03**
7	9.00	0.56	0.59	5.00	5.31	**1.06**	**0.44**	**2.55**
8	7.00	0.43	0.52	3.00	3.63	**1.21**	**0.39**	**3.75**
9	1.00	1.00	0.50	1.00	0.50	**0.50**	**0.07**	**3.55**
10	4.00	0.50	0.61	2.00	2.44	**1.22**	**0.30**	**4.87**
11	0.00	UtC	0.29	0.00	0.00	UtC	UtC	UtC
12	2.00	1.00	0.15	2.00	0.30	0.15	0.04	0.61
Total	78.00	0.68	0.62	53.00	48.44	**0.91**	**0.70**	**1.20**

Note. **Bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H36

DRAOR Total Score Absolute Predictive Accuracy for Any Return – Hispanic Men

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	22	0.45	0.43	10.00	9.44	**0.94**	**0.51**	**1.75**
Moderate	35	0.77	0.76	27.00	26.74	**0.99**	**0.68**	**1.44**
Mod.-High	21	0.76	0.86	16.00	18.12	**1.13**	**0.69**	**1.85**
High	0	UtC	1.00	0.00	0.00	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H37

DRAOR Stable Domain Score Absolute Predictive Accuracy for Technical Violations – White Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	40.00	0.28	0.17	11.00	6.96	**0.63**	**0.35**	**1.14**
1	81.00	0.47	0.34	38.00	27.86	**0.73**	**0.53**	**1.01**
2	126.00	0.51	0.52	64.00	65.39	**1.02**	**0.80**	**1.31**
3	214.00	0.57	0.51	123.00	109.35	**0.89**	**0.75**	**1.06**
4	197.00	0.66	0.56	131.00	111.11	**0.85**	**0.71**	**1.01**
5	259.00	0.66	0.61	170.00	156.70	**0.92**	**0.79**	**1.07**
6	302.00	0.71	0.68	215.00	204.15	**0.95**	**0.83**	**1.09**
7	288.00	0.68	0.78	195.00	224.06	**1.15**	**1.00**	**1.32**
8	243.00	0.74	0.69	179.00	168.16	**0.94**	**0.81**	**1.09**
9	194.00	0.73	0.71	141.00	136.96	**0.97**	**0.82**	**1.15**
10	117.00	0.79	1.00	92.00	117.00	1.27	1.04	1.56
11	70.00	0.64	1.00	45.00	70.00	1.56	1.16	2.08
12	36.00	0.64	1.00	23.00	36.00	1.57	1.04	2.36
Total	2167.00	0.66	0.56	1427.00	1207.02	**0.85**	**0.80**	**0.89**

Note. **Bold** denotes non-significant E/O index.

Table H38

DRAOR Acute Domain Score Absolute Predictive Accuracy for Technical Violations – White Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	13.00	0.31	0.08	4.00	1.08	0.27	0.10	0.72
1	39.00	0.44	0.23	17.00	9.01	0.53	0.33	0.85
2	74.00	0.36	0.39	27.00	28.79	**1.07**	**0.73**	**1.55**
3	134.00	0.57	0.50	77.00	67.00	**0.87**	**0.70**	**1.09**
4	169.00	0.61	0.56	103.00	93.80	**0.91**	**0.75**	**1.10**
5	200.00	0.62	0.59	124.00	117.60	**0.95**	**0.80**	**1.13**
6	246.00	0.64	0.68	158.00	166.05	**1.05**	**0.90**	**1.23**
7	262.00	0.66	0.64	172.00	166.63	**0.97**	**0.83**	**1.12**
8	267.00	0.72	0.77	191.00	204.26	**1.07**	**0.93**	**1.23**
9	265.00	0.65	0.73	171.00	194.25	**1.14**	**0.98**	**1.32**
10	195.00	0.77	0.80	151.00	156.00	**1.03**	**0.88**	**1.21**
11	155.00	0.78	0.90	121.00	139.50	**1.15**	**0.96**	**1.38**
12	89.00	0.79	1.00	70.00	89.00	1.27	1.01	1.61
13	41.00	0.68	1.00	28.00	41.00	1.46	1.01	2.12
14	18.00	0.72	1.00	13.00	18.00	**1.38**	**0.80**	**2.38**
Total	2167.00	0.66	0.56	1427.00	1207.02	0.85	0.80	0.89

Note. **Bold** denotes non-significant E/O index.

Table H39

DRAOR Protective Domain Score Absolute Predictive Accuracy for Technical Violations – White Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	65.00	0.68	1.00	44.00	65.00	1.48	1.10	1.99
1	81.00	0.79	0.93	64.00	75.57	**1.18**	**0.92**	**1.51**
2	112.00	0.70	0.88	78.00	98.00	1.26	1.01	1.57
3	160.00	0.66	0.90	106.00	144.00	1.36	1.12	1.64
4	193.00	0.73	0.68	141.00	131.24	**0.93**	**0.79**	**1.10**
5	232.00	0.70	0.59	162.00	137.58	0.85	0.73	0.99
6	513.00	0.69	0.54	352.00	278.05	0.79	0.71	0.88
7	233.00	0.69	0.51	160.00	119.53	0.75	0.64	0.87
8	157.00	0.62	0.48	97.00	75.52	0.78	0.64	0.95
9	138.00	0.58	0.44	80.00	60.44	0.76	0.61	0.94
10	110.00	0.55	0.39	61.00	43.01	0.71	0.55	0.91
11	78.00	0.50	0.21	39.00	16.22	0.42	0.30	0.57
12	95.00	0.45	0.12	43.00	11.50	0.27	0.20	0.36
Total	2167.00	0.66	0.56	1427.00	1207.02	0.85	0.80	0.89

Note. **bold** denotes non-significant E/O index.

Table H40

DRAOR Total Score Absolute Predictive Accuracy for Technical Violation – White Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	548	0.54	0.37	294.00	201.66	0.69	0.61	0.77
Moderate	760	0.67	0.69	511.00	520.60	**1.02**	**0.93**	**1.11**
Mod.-High	838	0.73	0.82	609.00	690.51	1.13	1.05	1.23
High	21	0.62	0.95	13.00	19.95	**1.53**	**0.89**	**2.64**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H41

DRAOR Stable Domain Score Absolute Predictive Accuracy for New Offence – White Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	40.00	0.05	0.13	2.00	5.20	**2.60**	**0.65**	**10.40**
1	81.00	0.12	0.19	10.00	15.23	**1.52**	**0.82**	**2.83**
2	126.00	0.09	0.37	11.00	46.62	4.24	2.35	7.65
3	214.00	0.07	0.34	15.00	72.76	4.85	2.92	8.05
4	197.00	0.11	0.27	22.00	53.78	2.44	1.61	3.71
5	259.00	0.10	0.28	25.00	72.26	2.89	1.95	4.28
6	302.00	0.12	0.35	37.00	106.30	2.87	2.08	3.97
7	288.00	0.12	0.37	34.00	106.56	3.13	2.24	4.39
8	243.00	0.12	0.31	29.00	74.84	2.58	1.79	3.71
9	194.00	0.14	0.35	27.00	68.48	2.54	1.74	3.70
10	117.00	0.11	0.20	13.00	23.40	1.80	1.05	3.10
11	70.00	0.10	0.50	7.00	35.00	5.00	2.38	10.49
12	36.00	0.14	0.34	5.00	12.24	2.45	1.02	5.88
Total	2167.00	0.11	0.30	237.00	656.60	2.77	2.44	3.15

Note. **bold** denotes non-significant E/O index.

Table H42

DRAOR Acute Domain Score Absolute Predictive Accuracy for New Offence – White Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	13.00	0.08	0.08	1.00	1.08	**1.08**	**0.15**	**7.66**
1	39.00	0.23	0.08	9.00	3.00	0.33	0.17	0.64
2	74.00	0.04	0.13	3.00	9.84	3.28	1.06	10.17
3	134.00	0.11	0.13	15.00	17.69	**1.18**	**0.71**	**1.96**
4	169.00	0.09	0.27	15.00	46.14	3.08	1.85	5.10
5	200.00	0.09	0.35	18.00	70.60	3.92	2.47	6.23
6	246.00	0.12	0.48	29.00	116.85	4.03	2.80	5.80
7	262.00	0.10	0.34	26.00	89.34	3.44	2.34	5.05
8	267.00	0.14	0.41	37.00	110.00	2.97	2.15	4.10
9	265.00	0.09	0.27	25.00	71.55	2.86	1.93	4.24
10	195.00	0.15	0.20	29.00	39.00	**1.34**	**0.93**	**1.94**
11	155.00	0.10	0.60	15.00	93.00	6.20	3.74	10.28
12	89.00	0.10	0.67	9.00	59.36	6.60	3.43	12.68
13	41.00	0.15	1.00	6.00	41.00	6.83	3.07	15.21
14	18.00	0.00	1.00	0.00	18.00	UtC	UtC	UtC
Total	2167.00	0.11	0.30	237.00	656.60	2.77	2.44	3.15

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H43

DRAOR Protective Domain Score Absolute Predictive Accuracy for New Offence – White Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	65.00	0.15	0.20	10.00	13.00	**1.30**	**0.70**	**2.42**
1	81.00	0.10	0.67	8.00	54.03	6.75	3.38	13.50
2	112.00	0.09	0.50	10.00	56.00	5.60	3.01	10.41
3	160.00	0.18	0.40	28.00	64.00	2.29	1.58	3.31
4	193.00	0.11	0.32	21.00	61.76	2.94	1.92	4.51
5	232.00	0.11	0.30	25.00	68.67	2.75	1.86	4.07
6	513.00	0.11	0.36	55.00	182.12	3.31	2.54	4.31
7	233.00	0.11	0.23	25.00	53.82	2.15	1.45	3.19
8	157.00	0.13	0.30	20.00	46.47	2.32	1.50	3.60
9	138.00	0.09	0.31	12.00	43.19	3.60	2.04	6.34
10	110.00	0.09	0.48	10.00	52.58	5.26	2.83	9.77
11	78.00	0.06	0.21	5.00	16.22	3.24	1.35	7.80
12	95.00	0.08	0.06	8.00	5.80	**0.72**	**0.36**	**1.45**
Total	2167.00	0.11	0.30	237.00	656.60	2.77	2.44	3.15

Note. **bold** denotes non-significant E/O index.

Table H44

DRAOR Total Score Absolute Predictive Accuracy for New Offence – White Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	548	0.10	0.23	56.00	127.68	2.28	1.75	2.96
Moderate	760	0.09	0.38	72.00	287.28	3.99	3.17	5.03
Mod.-High	838	0.13	0.33	108.00	279.05	2.58	2.14	3.12
High	21	0.05	0.50	1.00	10.50	10.50	1.48	74.54

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H45

DRAOR Stable Domain Score Absolute Predictive Accuracy for Any Return – White Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	40.00	0.30	0.22	12.00	8.68	**0.72**	**0.41**	**1.27**
1	81.00	0.49	0.34	40.00	27.86	0.70	0.51	0.95
2	126.00	0.51	0.59	64.00	74.72	**1.17**	**0.91**	**1.49**
3	214.00	0.58	0.62	125.00	132.04	**1.06**	**0.89**	**1.26**
4	197.00	0.69	0.49	135.00	95.55	0.71	0.60	0.84
5	259.00	0.68	0.65	175.00	168.61	**0.96**	**0.83**	**1.12**
6	302.00	0.74	0.75	222.00	225.29	**1.01**	**0.89**	**1.16**
7	288.00	0.69	0.85	200.00	245.38	1.23	1.07	1.41
8	243.00	0.74	0.69	181.00	168.16	**0.93**	**0.80**	**1.07**
9	194.00	0.75	0.77	146.00	148.41	**1.02**	**0.86**	**1.20**
10	117.00	0.80	1.00	94.00	117.00	1.24	1.02	1.52
11	70.00	0.67	1.00	47.00	70.00	1.49	1.12	1.98
12	36.00	0.64	1.00	23.00	36.00	1.57	1.04	2.36
Total	2167.00	0.68	0.62	1464.00	1345.71	0.92	0.87	0.97

Note. **bold** denotes non-significant E/O index.

Table H46

DRAOR Acute Domain Score Absolute Predictive Accuracy for Any Return – White Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	13.00	0.31	0.17	4.00	2.17	**0.54**	**0.20**	**1.45**
1	39.00	0.51	0.27	20.00	10.49	0.52	0.34	0.81
2	74.00	0.38	0.44	28.00	32.86	**1.17**	**0.81**	**1.70**
3	134.00	0.60	0.50	81.00	67.00	**0.83**	**0.67**	**1.03**
4	169.00	0.62	0.50	104.00	84.50	0.81	0.67	0.98
5	200.00	0.63	0.68	126.00	135.20	**1.07**	**0.90**	**1.28**
6	246.00	0.65	0.80	160.00	196.80	1.23	1.05	1.44
7	262.00	0.67	0.75	175.00	196.50	**1.12**	**0.97**	**1.30**
8	267.00	0.75	0.85	201.00	227.75	**1.13**	**0.99**	**1.30**
9	265.00	0.67	0.73	177.00	194.25	**1.10**	**0.95**	**1.27**
10	195.00	0.79	1.00	154.00	195.00	1.27	1.08	1.48
11	155.00	0.79	0.80	123.00	124.00	**1.01**	**0.84**	**1.20**
12	89.00	0.79	1.00	70.00	89.00	1.27	1.01	1.61
13	41.00	0.68	1.00	28.00	41.00	1.46	1.01	2.12
14	18.00	0.72	1.00	13.00	18.00	**1.38**	**0.80**	**2.38**
Total	2167.00	0.68	0.62	1464.00	1345.71	0.92	0.87	0.97

Note. **bold** denotes non-significant E/O index.

Table H47

DRAOR Protective Domain Score Absolute Predictive Accuracy for Any Return – White Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	65.00	0.69	1.00	45.00	65.00	1.44	1.08	1.93
1	81.00	0.79	0.67	64.00	54.03	**0.84**	**0.66**	**1.08**
2	112.00	0.71	0.88	80.00	98.00	**1.23**	**0.98**	**1.53**
3	160.00	0.69	0.90	111.00	144.00	1.30	1.08	1.56
4	193.00	0.74	0.72	142.00	138.96	**0.98**	**0.83**	**1.15**
5	232.00	0.72	0.67	168.00	154.74	**0.92**	**0.79**	**1.07**
6	513.00	0.71	0.80	364.00	408.35	1.12	1.01	1.24
7	233.00	0.70	0.59	162.00	137.47	0.85	0.73	0.99
8	157.00	0.64	0.52	100.00	81.48	0.81	0.67	0.99
9	138.00	0.59	0.50	82.00	69.00	**0.84**	**0.68**	**1.04**
10	110.00	0.56	0.61	62.00	66.99	**1.08**	**0.84**	**1.39**
11	78.00	0.53	0.29	41.00	22.78	0.56	0.41	0.75
12	95.00	0.45	0.15	43.00	14.44	0.34	0.25	0.45
Total	2167.00	0.68	0.62	1464.00	1345.71	0.92	0.87	0.97

Note. **bold** denotes non-significant E/O index.

Table H48

DRAOR Total Score Absolute Predictive Accuracy for Any Return – White Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	548	0.55	0.43	303.00	235.09	0.78	0.69	0.87
Moderate	760	0.69	0.76	523.00	580.64	1.11	1.02	1.21
Mod.-High	838	0.75	0.86	625.00	723.19	1.16	1.07	1.25
High	21	0.62	1.00	13.00	21.00	**1.62**	**0.94**	**2.78**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H49

DRAOR Stable Domain Score Absolute Predictive Accuracy for Technical Violations – Black Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	14.00	0.64	0.17	9.00	2.44	0.27	0.14	0.52
1	12.00	0.92	0.34	11.00	4.13	0.38	0.21	0.68
2	35.00	0.71	0.52	25.00	18.17	**0.73**	**0.49**	**1.08**
3	45.00	0.82	0.51	37.00	23.00	0.62	0.45	0.86
4	71.00	0.79	0.56	56.00	40.04	0.72	0.55	0.93
5	67.00	0.73	0.61	49.00	40.54	**0.83**	**0.63**	**1.09**
6	86.00	0.79	0.68	68.00	58.14	**0.85**	**0.67**	**1.08**
7	90.00	0.83	0.78	75.00	70.02	**0.93**	**0.74**	**1.17**
8	65.00	0.82	0.69	53.00	44.98	**0.85**	**0.65**	**1.11**
9	64.00	0.81	0.71	52.00	45.18	**0.87**	**0.66**	**1.14**
10	40.00	0.78	1.00	31.00	40.00	**1.29**	**0.91**	**1.83**
11	15.00	0.87	1.00	13.00	15.00	**1.15**	**0.67**	**1.99**
12	20.00	0.85	1.00	17.00	20.00	**1.18**	**0.73**	**1.89**
Total	496.00	0.74	0.56	368.00	276.27	0.75	0.68	0.83

Note. **bold** denotes non-significant E/O index.

Table H50

DRAOR Acute Domain Score Absolute Predictive Accuracy for Technical Violations – Black Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	5.00	0.40	0.08	2.00	0.42	0.21	0.05	0.83
1	8.00	0.63	0.23	5.00	1.85	0.37	0.15	0.89
2	13.00	0.62	0.39	8.00	5.06	**0.63**	**0.32**	**1.26**
3	28.00	0.71	0.50	20.00	14.00	**0.70**	**0.45**	**1.09**
4	44.00	0.75	0.56	33.00	24.42	**0.74**	**0.53**	**1.04**
5	68.00	0.79	0.59	54.00	39.98	0.74	0.57	0.97
6	68.00	0.72	0.68	49.00	45.90	**0.94**	**0.71**	**1.24**
7	74.00	0.77	0.64	57.00	47.06	**0.83**	**0.64**	**1.07**
8	63.00	0.67	0.77	42.00	48.20	**1.15**	**0.85**	**1.55**
9	39.00	0.69	0.73	27.00	28.59	**1.06**	**0.73**	**1.54**
10	29.00	0.83	0.80	24.00	23.20	**0.97**	**0.65**	**1.44**
11	30.00	0.77	0.90	23.00	27.00	**1.17**	**0.78**	**1.77**
12	18.00	0.89	1.00	16.00	18.00	**1.13**	**0.69**	**1.84**
13	5.00	1.00	1.00	5.00	5.00	**1.00**	**0.42**	**2.40**
14	4.00	0.75	1.00	3.00	4.00	**1.33**	**0.43**	**4.13**
Total	496.00	0.74	0.56	368.00	276.27	0.75	0.68	0.83

Note. **bold** denotes non-significant E/O index.

Table H51

DRAOR Protective Domain Score Absolute Predictive Accuracy for Technical Violations – Black Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	18.00	0.67	1.00	12.00	18.00	**1.50**	**0.85**	**2.64**
1	15.00	0.80	0.93	12.00	14.00	**1.17**	**0.66**	**2.05**
2	27.00	0.67	0.88	18.00	23.63	**1.31**	**0.83**	**2.08**
3	32.00	0.78	0.90	25.00	28.80	**1.15**	**0.78**	**1.70**
4	46.00	0.76	0.68	35.00	31.28	**0.89**	**0.64**	**1.24**
5	60.00	0.82	0.59	49.00	35.58	0.73	0.55	0.96
6	135.00	0.73	0.54	99.00	73.17	0.74	0.61	0.90
7	49.00	0.78	0.51	38.00	25.14	0.66	0.48	0.91
8	36.00	0.81	0.48	29.00	17.32	0.60	0.41	0.86
9	32.00	0.69	0.44	22.00	14.02	0.64	0.42	0.97
10	18.00	0.72	0.39	13.00	7.04	0.54	0.31	0.93
11	14.00	0.57	0.21	8.00	2.91	0.36	0.18	0.73
12	14.00	0.57	0.12	8.00	1.69	0.21	0.11	0.42
Total	496.00	0.74	0.56	368.00	276.27	0.75	0.68	0.83

Note. **bold** denotes non-significant E/O index.

Table H52

DRAOR Total Score Absolute Predictive Accuracy for Technical Violation – Black Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	109.00	0.72	0.37	78.00	40.11	0.51	0.41	0.64
Moderate	212.00	0.74	0.69	156.00	145.22	**0.93**	**0.80**	**1.09**
Mod.-High	168.00	0.76	0.82	128.00	138.43	**1.08**	**0.91**	**1.29**
High	7.00	0.86	0.95	6.00	6.65	**1.11**	**0.50**	**2.47**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H53

DRAOR Stable Domain Score Absolute Predictive Accuracy for New Offence – Black Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	9.00	0.00	0.13	0.00	1.17	UtC	UtC	UtC
1	11.00	0.00	0.19	0.00	2.07	UtC	UtC	UtC
2	25.00	0.08	0.37	2.00	9.25	4.63	1.16	18.49
3	37.00	0.08	0.34	3.00	12.58	4.19	1.35	13.00
4	56.00	0.11	0.27	6.00	15.29	2.55	1.14	5.67
5	49.00	0.06	0.28	3.00	13.67	4.56	1.47	14.13
6	68.00	0.06	0.35	4.00	23.94	5.98	2.25	15.94
7	75.00	0.13	0.37	10.00	27.75	2.78	1.49	5.16
8	53.00	0.19	0.31	10.00	16.32	**1.63**	**0.88**	**3.03**
9	52.00	0.10	0.35	5.00	18.36	3.67	1.53	8.82
10	31.00	0.10	0.20	3.00	6.20	**2.07**	**0.67**	**6.41**
11	13.00	0.23	0.50	3.00	6.50	**2.17**	**0.70**	**6.72**
12	17.00	0.18	0.34	3.00	5.78	**1.93**	**0.62**	**5.97**
Total	496.00	0.10	0.30	52.00	150.29	2.89	2.20	3.79

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H54

DRAOR Acute Domain Score Absolute Predictive Accuracy for New Offence – Black Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	5.00	0.00	0.08	0.00	0.42	UtC	UtC	UtC
1	8.00	0.00	0.08	0.00	0.62	UtC	UtC	UtC
2	13.00	0.08	0.13	1.00	1.73	**1.73**	**0.24**	**12.27**
3	28.00	0.00	0.13	0.00	3.70	UtC	UtC	UtC
4	44.00	0.07	0.27	3.00	12.01	4.00	1.29	12.41
5	68.00	0.12	0.35	8.00	24.00	3.00	1.50	6.00
6	68.00	0.07	0.48	5.00	32.30	6.46	2.69	15.52
7	74.00	0.16	0.34	12.00	25.23	2.10	1.19	3.70
8	63.00	0.16	0.41	10.00	25.96	2.60	1.40	4.82
9	39.00	0.05	0.27	2.00	10.53	5.27	1.32	21.05
10	29.00	0.14	0.20	4.00	5.80	**1.45**	**0.54**	**3.86**
11	30.00	0.10	0.60	3.00	18.00	6.00	1.94	18.60
12	18.00	0.17	0.67	3.00	12.01	4.00	1.29	12.41
13	5.00	0.00	1.00	0.00	5.00	UtC	UtC	UtC
14	4.00	0.25	1.00	1.00	4.00	**4.00**	**0.56**	**28.40**
Total	496.00	0.10	0.30	52.00	150.29	2.89	2.20	3.79

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H55

DRAOR Protective Domain Score Absolute Predictive Accuracy for New Offence – Black Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	18	0.11	0.20	2.00	3.60	**1.80**	**0.45**	**7.20**
1	15	0.20	0.67	3.00	10.01	3.34	1.08	10.34
2	27	0.07	0.50	2.00	13.50	6.75	1.69	26.99
3	32	0.13	0.40	4.00	12.80	3.20	1.20	8.53
4	46	0.13	0.32	6.00	14.72	2.45	1.10	5.46
5	60	0.10	0.30	6.00	17.76	2.96	1.33	6.59
6	135	0.12	0.36	16.00	47.93	3.00	1.84	4.89
7	49	0.04	0.23	2.00	11.32	5.66	1.42	22.63
8	36	0.11	0.30	4.00	10.66	2.66	1.00	7.10
9	32	0.06	0.31	2.00	10.02	5.01	1.25	20.02
10	18	0.11	0.48	2.00	8.60	4.30	1.08	17.20
11	14	0.00	0.21	0.00	2.91	UtC	UtC	UtC
12	14	0.21	0.06	3.00	0.85	0.28	0.09	0.88
Total	496	0.10	0.30	52.00	150.29	2.89	2.20	3.79

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H56

DRAOR Total Score Absolute Predictive Accuracy for New Offence –Black Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	109	0.06	0.23	6.00	25.40	4.23	1.90	9.42
Moderate	212	0.11	0.38	24.00	80.14	3.34	2.24	4.98
Mod.-High	168	0.13	0.33	22.00	55.94	2.54	1.67	3.86
High	7	0.00	0.50	0.00	3.50	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H57

DRAOR Stable Domain Score Absolute Predictive Accuracy for Any Return – Black Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	9.00	0.44	0.22	4.00	1.95	UtC	UtC	UtC
1	11.00	0.91	0.34	10.00	3.78	0.38	0.20	0.70
2	25.00	0.64	0.59	16.00	14.83	**0.93**	**0.57**	**1.51**
3	37.00	0.78	0.62	29.00	22.83	**0.79**	**0.55**	**1.13**
4	56.00	0.77	0.49	43.00	27.16	0.63	0.47	0.85
5	49.00	0.63	0.65	31.00	31.90	**1.03**	**0.72**	**1.46**
6	68.00	0.74	0.75	50.00	50.73	**1.01**	**0.77**	**1.34**
7	75.00	0.81	0.85	61.00	63.90	**1.05**	**0.82**	**1.35**
8	53.00	0.81	0.69	43.00	36.68	**0.85**	**0.63**	**1.15**
9	52.00	0.79	0.77	41.00	39.78	**0.97**	**0.71**	**1.32**
10	31.00	0.74	1.00	23.00	31.00	**1.35**	**0.90**	**2.03**
11	13.00	0.85	1.00	11.00	13.00	**1.18**	**0.65**	**2.13**
12	17.00	0.82	1.00	14.00	17.00	**1.21**	**0.72**	**2.05**
Total	496.00	0.76	0.62	376.00	308.02	0.82	0.74	0.91

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H58

DRAOR Acute Domain Score Absolute Predictive Accuracy for Any Return – Black Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	5.00	0.40	0.17	2.00	0.84	**0.42**	**0.10**	**1.67**
1	8.00	0.63	0.27	5.00	2.15	**0.43**	**0.18**	**1.03**
2	13.00	0.62	0.44	8.00	5.77	**0.72**	**0.36**	**1.44**
3	28.00	0.71	0.50	20.00	14.00	**0.70**	**0.45**	**1.09**
4	44.00	0.75	0.50	33.00	22.00	0.67	0.47	0.94
5	68.00	0.81	0.68	55.00	45.97	**0.84**	**0.64**	**1.09**
6	68.00	0.72	0.80	49.00	54.40	**1.11**	**0.84**	**1.47**
7	74.00	0.81	0.75	60.00	55.50	**0.93**	**0.72**	**1.19**
8	63.00	0.70	0.85	44.00	53.74	**1.22**	**0.91**	**1.64**
9	39.00	0.69	0.73	27.00	28.59	**1.06**	**0.73**	**1.54**
10	29.00	0.86	1.00	25.00	29.00	**1.16**	**0.78**	**1.72**
11	30.00	0.80	0.80	24.00	24.00	**1.00**	**0.67**	**1.49**
12	18.00	0.89	1.00	16.00	18.00	**1.13**	**0.69**	**1.84**
13	5.00	1.00	1.00	5.00	5.00	**1.00**	**0.42**	**2.40**
14	4.00	0.75	1.00	3.00	4.00	**1.33**	**0.43**	**4.13**
Total	496.00	0.76	0.62	376.00	308.02	0.82	0.74	0.91

Note. **bold** denotes non-significant E/O index.

Table H59

DRAOR Protective Domain Score Absolute Predictive Accuracy for Any Return – Black Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	18.00	0.67	1.00	12.00	18.00	**1.50**	**0.85**	**2.64**
1	15.00	0.87	0.67	13.00	10.01	**0.77**	**0.45**	**1.33**
2	27.00	0.67	0.88	18.00	23.63	**1.31**	**0.83**	**2.08**
3	32.00	0.78	0.90	25.00	28.80	**1.15**	**0.78**	**1.70**
4	46.00	0.78	0.72	36.00	33.12	**0.92**	**0.66**	**1.28**
5	60.00	0.82	0.67	49.00	40.02	**0.82**	**0.62**	**1.08**
6	135.00	0.76	0.80	102.00	107.46	**1.05**	**0.87**	**1.28**
7	49.00	0.80	0.59	39.00	28.91	**0.74**	**0.54**	**1.01**
8	36.00	0.81	0.52	29.00	18.68	0.64	0.45	0.93
9	32.00	0.72	0.50	23.00	16.00	**0.70**	**0.46**	**1.05**
10	18.00	0.72	0.61	13.00	10.96	**0.84**	**0.49**	**1.45**
11	14.00	0.57	0.29	8.00	4.09	**0.51**	**0.26**	**1.02**
12	14.00	0.64	0.15	9.00	2.13	0.24	0.12	0.45
Total	496.00	0.76	0.62	376.00	308.02	0.82	0.74	0.91

Note. **bold** denotes non-significant E/O index.

Table H60

DRAOR Total Score Absolute Predictive Accuracy for Any Return – Black Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	109	0.72	0.43	79.00	46.76	0.59	0.47	0.74
Moderate	212	0.75	0.76	160.00	161.97	**1.01**	**0.87**	**1.18**
Mod.-High	168	0.78	0.86	131.00	144.98	**1.11**	**0.93**	**1.31**
High	7	0.86	1.00	6.00	7.00	**1.17**	**0.52**	**2.60**

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index.

Table H61

DRAOR Stable Domain Score Absolute Predictive Accuracy for Technical Violations – Hispanic Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0			0.17	0.00	0.00	UtC	UtC	UtC
1	5.00	0.80	0.34	4.00	1.72	**0.43**	**0.16**	**1.15**
2	8.00	0.38	0.52	3.00	4.15	**1.38**	**0.45**	**4.29**
3	5.00	0.20	0.51	1.00	2.56	**2.56**	**0.36**	**18.14**
4	13.00	0.38	0.56	5.00	7.33	**1.47**	**0.61**	**3.52**
5	11.00	0.82	0.61	9.00	6.66	**0.74**	**0.38**	**1.42**
6	15.00	0.80	0.68	12.00	10.14	**0.85**	**0.48**	**1.49**
7	9.00	0.56	0.78	5.00	7.00	**1.40**	**0.58**	**3.36**
8	12.00	0.75	0.69	9.00	8.30	**0.92**	**0.48**	**1.77**
9	11.00	0.91	0.71	10.00	7.77	**0.78**	**0.42**	**1.44**
10	6.00	1.00	1.00	6.00	6.00	**1.00**	**0.45**	**2.23**
11	3.00	0.00	1.00	0.00	3.00	UtC	UtC	UtC
12	2.00	0.50	1.00	1.00	2.00	**2.00**	**0.28**	**14.20**
Total	100.00	0.65	0.56	65.00	55.70	**0.86**	**0.67**	**1.09**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H62

DRAOR Acute Domain Score Absolute Predictive Accuracy for Technical Violations – Hispanic Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	0.00	UtC	UtC	UtC	0.00	UtC	UtC	UtC
1	4.00	0.75	0.23	3.00	0.92	0.31	0.10	0.95
2	7.00	0.43	0.39	3.00	2.72	**0.91**	**0.29**	**2.81**
3	4.00	0.50	0.50	2.00	2.00	**1.00**	**0.25**	**4.00**
4	7.00	0.43	0.56	3.00	3.89	**1.30**	**0.42**	**4.02**
5	8.00	0.75	0.59	6.00	4.70	**0.78**	**0.35**	**1.75**
6	15.00	0.67	0.68	10.00	10.13	**1.01**	**0.54**	**1.88**
7	11.00	0.64	0.64	7.00	7.00	**1.00**	**0.48**	**2.10**
8	12.00	0.75	0.77	9.00	9.18	**1.02**	**0.53**	**1.96**
9	10.00	0.80	0.73	8.00	7.33	**0.92**	**0.46**	**1.83**
10	10.00	0.80	0.80	8.00	8.00	**1.00**	**0.50**	**2.00**
11	7.00	0.57	0.90	4.00	6.30	**1.58**	**0.59**	**4.20**
12	4.00	0.25	1.00	1.00	4.00	**4.00**	**0.56**	**28.40**
13	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
14	0.00	0.00	1.00	0.00	0.00	UtC	UtC	UtC
Total	100.00	0.65	0.56	65.00	55.70	**0.86**	**0.67**	**1.09**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H63

DRAOR Protective Domain Score Absolute Predictive Accuracy for Technical Violations – Hispanic Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	4.00	0.50	1.00	2.00	4.00	**2.00**	**0.50**	**8.00**
1	5.00	0.60	0.93	3.00	4.67	**1.56**	**0.50**	**4.82**
2	6.00	0.33	0.88	2.00	5.25	**2.63**	**0.66**	**10.50**
3	8.00	1.00	0.90	8.00	7.20	**0.90**	**0.45**	**1.80**
4	3.00	0.67	0.68	2.00	2.04	**1.02**	**0.26**	**4.08**
5	12.00	0.67	0.59	8.00	7.12	**0.89**	**0.44**	**1.78**
6	30.00	0.73	0.54	22.00	16.26	**0.74**	**0.49**	**1.12**
7	11.00	0.55	0.51	6.00	5.64	**0.94**	**0.42**	**2.09**
8	4.00	1.00	0.48	4.00	1.92	**0.48**	**0.18**	**1.28**
9	5.00	0.20	0.44	1.00	2.19	**2.19**	**0.31**	**15.55**
10	5.00	0.40	0.39	2.00	1.96	**0.98**	**0.24**	**3.91**
11	6.00	0.83	0.21	5.00	1.25	0.25	0.10	0.60
12	1.00	0.00	0.12	0.00	0.12	UtC	UtC	UtC
Total	100.00	0.65	0.56	65.00	55.70	**0.86**	**0.67**	**1.09**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H64

DRAOR Total Score Absolute Predictive Accuracy for Technical Violations – Hispanic Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	23	0.57	0.37	13.00	8.46	**0.65**	**0.38**	**1.12**
Moderate	40	0.63	0.69	25.00	27.40	**1.10**	**0.74**	**1.62**
Mod.-High	36	0.75	0.82	27.00	29.66	**1.10**	**0.75**	**1.60**
High	1	0.00	0.95	0.00	0.95	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H65

DRAOR Stable Domain Score Absolute Predictive Accuracy for New Offence – Hispanic Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	0.00	UtC	0.13	0.00	0.00	UtC	UtC	UtC
1	5.00	0.00	0.19	0.00	0.94	UtC	UtC	UtC
2	8.00	0.00	0.37	0.00	2.96	UtC	UtC	UtC
3	5.00	0.00	0.34	0.00	1.70	UtC	UtC	UtC
4	13.00	0.00	0.27	0.00	3.55	UtC	UtC	UtC
5	11.00	0.18	0.28	2.00	3.07	**1.53**	**0.38**	**6.14**
6	15.00	0.07	0.35	1.00	5.28	**5.28**	**0.74**	**37.48**
7	9.00	0.22	0.37	2.00	3.33	**1.67**	**0.42**	**6.66**
8	12.00	0.00	0.31	0.00	3.70	UtC	UtC	UtC
9	11.00	0.09	0.35	1.00	3.88	**3.88**	**0.55**	**27.57**
10	6.00	0.00	0.20	0.00	1.20	UtC	UtC	UtC
11	3.00	0.00	0.50	0.00	1.50	UtC	UtC	UtC
12	2.00	0.00	0.34	0.00	0.68	UtC	UtC	UtC
Total	100.00	0.06	0.30	6.00	30.30	5.05	2.27	11.24

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H66

DRAOR Acute Domain Score Absolute Predictive Accuracy for New Offence – Hispanic Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	0.00	UtC	0.08	0.00	0.00	UtC	UtC	UtC
1	4.00	0.25	0.08	1.00	0.31	**0.31**	**0.04**	**2.19**
2	7.00	0.00	0.13	0.00	0.93	UtC	UtC	UtC
3	4.00	0.00	0.13	0.00	0.53	UtC	UtC	UtC
4	7.00	0.00	0.27	0.00	1.91	UtC	UtC	UtC
5	8.00	0.00	0.35	0.00	2.82	UtC	UtC	UtC
6	15.00	0.13	0.48	2.00	7.13	**3.56**	**0.89**	**14.24**
7	11.00	0.18	0.34	2.00	3.75	**1.88**	**0.47**	**7.50**
8	12.00	0.08	0.41	1.00	4.94	**4.94**	**0.70**	**35.10**
9	10.00	0.00	0.27	0.00	2.70	UtC	UtC	UtC
10	10.00	0.00	0.20	0.00	2.00	UtC	UtC	UtC
11	7.00	0.00	0.60	0.00	4.20	UtC	UtC	UtC
12	4.00	0.00	0.67	0.00	2.67	UtC	UtC	UtC
13	1.00	0.00	1.00	0.00	1.00	UtC	UtC	UtC
14	0.00	UtC	1.00	0.00	0.00	UtC	UtC	UtC
Total	100.00	0.06	0.30	6.00	30.30	**5.05**	**2.27**	**11.24**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H67

DRAOR Protective Domain Score Absolute Predictive Accuracy for New Offence – Hispanic Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	4.00	0.00	0.20	0.00	0.80	UtC	UtC	UtC
1	5.00	0.00	0.67	0.00	3.34	UtC	UtC	UtC
2	6.00	0.00	0.50	0.00	3.00	UtC	UtC	UtC
3	8.00	0.13	0.40	1.00	3.20	**3.20**	**0.45**	**22.72**
4	3.00	0.33	0.32	1.00	0.96	**0.96**	**0.14**	**6.82**
5	12.00	0.00	0.30	0.00	3.55	UtC	UtC	UtC
6	30.00	0.10	0.36	3.00	10.65	3.55	1.14	11.01
7	11.00	0.09	0.23	1.00	2.54	**2.54**	**0.36**	**18.04**
8	4.00	0.00	0.30	0.00	1.18	UtC	UtC	UtC
9	5.00	0.00	0.31	0.00	1.57	UtC	UtC	UtC
10	5.00	0.00	0.48	0.00	2.39	UtC	UtC	UtC
11	6.00	0.00	0.21	0.00	1.25	UtC	UtC	UtC
12	1.00	0.00	0.06	0.00	0.06	UtC	UtC	UtC
Total	100.00	0.06	0.30	6.00	30.30	5.05	2.27	11.24

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H68

DRAOR Total Score Absolute Predictive Accuracy for New Offence – Hispanic Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	23	0.00	0.23	0.00	5.36	UtC	UtC	UtC
Moderate	40	0.13	0.38	5.00	15.12	3.02	1.26	7.27
Mod.-High	36	0.03	0.33	1.00	11.99	11.99	1.69	85.11
High	1	0.00	0.50	0.00	0.50	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H69

DRAOR Stable Domain Score Absolute Predictive Accuracy for Any Return – Hispanic Women

Stable domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	0.00	UtC	0.22	0.00	0.00	UtC	UtC	UtC
1	5.00	0.80	0.34	4.00	1.72	**0.43**	**0.16**	**1.15**
2	8.00	0.38	0.59	3.00	4.74	**1.58**	**0.51**	**4.90**
3	5.00	0.20	0.62	1.00	3.09	**3.09**	**0.43**	**21.90**
4	13.00	0.38	0.49	5.00	6.31	**1.26**	**0.52**	**3.03**
5	11.00	0.82	0.65	9.00	7.16	**0.80**	**0.41**	**1.53**
6	15.00	0.80	0.75	12.00	11.19	**0.93**	**0.53**	**1.64**
7	9.00	0.78	0.85	7.00	7.67	**1.10**	**0.52**	**2.30**
8	11.00	0.82	0.69	9.00	7.61	**0.85**	**0.44**	**1.63**
9	11.00	0.91	0.77	10.00	8.42	**0.84**	**0.45**	**1.56**
10	6.00	1.00	1.00	6.00	6.00	**1.00**	**0.45**	**2.23**
11	3.00	0.00	1.00	0.00	3.00	UtC	UtC	UtC
12	2.00	0.50	1.00	1.00	2.00	2.00	0.28	14.20
Total	99.00	0.68	0.62	67.00	61.48	**0.92**	**0.72**	**1.17**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H70

DRAOR Acute Domain Score Absolute Predictive Accuracy for Any Return – Hispanic Women

Acute domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	0.00	UtC	0.17	0.00	0.00	UtC	UtC	UtC
1	4.00	1.00	0.27	4.00	1.08	0.27	0.10	0.72
2	7.00	0.43	0.44	3.00	3.11	**1.04**	**0.33**	**3.21**
3	4.00	0.50	0.50	2.00	2.00	**1.00**	**0.25**	**4.00**
4	7.00	0.43	0.50	3.00	3.50	**1.17**	**0.38**	**3.62**
5	8.00	0.75	0.68	6.00	5.41	**0.90**	**0.40**	**2.01**
6	15.00	0.73	0.80	11.00	12.00	**1.09**	**0.60**	**1.97**
7	11.00	0.64	0.75	7.00	8.25	**1.18**	**0.56**	**2.47**
8	12.00	0.75	0.85	9.00	10.24	**1.14**	**0.59**	**2.19**
9	10.00	0.80	0.73	8.00	7.33	**0.92**	**0.46**	**1.83**
10	10.00	0.80	1.00	8.00	10.00	**1.25**	**0.63**	**2.50**
11	6.00	0.67	0.80	4.00	4.80	**1.20**	**0.45**	**3.20**
12	4.00	0.25	1.00	1.00	4.00	**4.00**	**0.56**	**28.40**
13	1.00	1.00	1.00	1.00	1.00	**1.00**	**0.14**	**7.10**
14	0.00	UtC	1.00	0.00	0.00	UtC	UtC	UtC
Total	99.00	0.68	0.62	67.00	61.48	**0.92**	**0.72**	**1.17**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H71

DRAOR Protective Domain Score Absolute Predictive Accuracy for Any Return – Hispanic Women

Protective domain Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
0	4	0.50	1.00	2.00	4.00	**2.00**	**0.50**	**8.00**
1	4	0.75	0.67	3.00	2.67	**0.89**	**0.29**	**2.76**
2	6	0.33	0.88	2.00	5.25	**2.63**	**0.66**	**10.50**
3	8	1.00	0.90	8.00	7.20	**0.90**	**0.45**	**1.80**
4	3	1.00	0.72	3.00	2.16	**0.72**	**0.23**	**2.23**
5	12	0.67	0.67	8.00	8.00	**1.00**	**0.50**	**2.00**
6	30	0.73	0.80	22.00	23.88	**1.09**	**0.71**	**1.65**
7	11	0.64	0.59	7.00	6.49	**0.93**	**0.44**	**1.94**
8	4	1.00	0.52	4.00	2.08	**0.52**	**0.19**	**1.38**
9	5	0.20	0.50	1.00	2.50	**2.50**	**0.35**	**17.75**
10	5	0.40	0.61	2.00	3.05	**1.52**	**0.38**	**6.09**
11	6	0.83	0.29	5.00	1.75	0.35	0.15	0.84
12	1	0.00	0.15	0.00	0.15	UtC	UtC	UtC
Total	99	0.68	0.62	67.00	61.48	**0.92**	**0.72**	**1.17**

Note. **bold** denotes non-significant E/O index, UtC = unable to calculate.

Table H72

DRAOR Total Score Absolute Predictive Accuracy for Any Return – Hispanic Women

DRAOR Total Score	*N*	Observed probability	Predicted probability	Observed recidivists (O)	Expected recidivists (E)	E/O Index	95% CI Lower Bound	95% CI Upper Bound
Low-Mod.	23	0.57	0.43	13.00	9.87	**0.76**	**0.44**	**1.31**
Moderate	40	0.68	0.76	27.00	30.56	**1.13**	**0.78**	**1.65**
Mod.-High	36	0.75	0.86	27.00	31.07	**1.15**	**0.79**	**1.68**
High	1	0.00	1.00	0.00	1.00	UtC	UtC	UtC

Note. Low-Mod. = Low-Moderate, Mod.-High = Moderate-High risk, **bold** denotes non-significant E/O index, UtC = unable to calculate.

Appendix I: Logistic Regression Item-Level Results

Table I1

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Technical Violations for White Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.141	.080	3.124	1	.077	1.152	.985	1.347
Attitudes towards Authority	-.034	.078	.189	1	.664	.967	.830	1.126
Impulse Control	.076	.088	.750	1	.386	1.079	.908	1.283
Problem Solving	.153	.089	2.972	1	.085	1.165	.979	1.386
Sense of Entitlement	-.012	.074	.029	1	.866	.988	.854	1.141
Attachment with Others	-.093	.081	1.307	1	.253	.911	.777	1.069
Substance Abuse	.187	.064	8.525	1	**.004**	1.206	1.063	1.367
Anger	.043	.069	.379	1	.538	1.044	.911	1.195
Access to Victims	-.062	.066	.871	1	.351	.940	.825	1.071
Negative Mood	.072	.065	1.232	1	.267	1.075	.946	1.220
Employment	.108	.058	3.547	1	.060	1.114	.996	1.248
Interpersonal Relationships	.324	.067	23.077	1	**<.001***	1.382	1.211	1.578
Living Situation	-.213	.067	10.172	1	**.001***	.808	.709	.921
Response to advice	-.082	.092	.795	1	.372	.922	.770	1.103
Prosocial Identity	-.115	.100	1.330	1	.249	.891	.733	1.084
High Expectations	-.135	.091	2.211	1	.137	.873	.731	1.044
Costs/Benefits	.038	.094	.165	1	.684	1.039	.864	1.248
Social Support	.056	.087	.422	1	.516	1.058	.893	1.254
Social Control	.026	.096	.073	1	.787	1.026	.850	1.239

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I2

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Technical Violations for Black Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.250	.170	2.160	1	.142	1.284	.920	1.792
Attitudes towards Authority	-.151	.170	.792	1	.373	.860	.616	1.199
Impulse Control	-.202	.193	1.087	1	.297	.817	.560	1.194
Problem Solving	.236	.199	1.404	1	.236	1.266	.857	1.869
Sense of Entitlement	.299	.172	3.021	1	.082	1.348	.963	1.888
Attachment with Others	-.164	.177	.862	1	.353	.848	.600	1.200
Substance Abuse	.268	.135	3.935	1	**.047**	1.307	1.003	1.702
Anger	-.157	.156	1.014	1	.314	.854	.629	1.161
Access to Victims	-.098	.142	.479	1	.489	.906	.686	1.197
Negative Mood	.011	.150	.005	1	.944	1.011	.753	1.356
Employment	-.078	.127	.380	1	.537	.925	.721	1.186
Interpersonal Relationships	.627	.160	15.367	1	**<.001***	1.873	1.369	2.563
Living Situation	-.599	.141	17.945	1	**<.001***	.549	.416	.725
Response to advice	-.247	.200	1.533	1	.216	.781	.528	1.155
Prosocial Identity	.323	.220	2.146	1	.143	1.381	.897	2.127
High Expectations	-.635	.205	9.559	1	**.002***	.530	.355	.793
Costs/Benefits	.133	.208	.410	1	.522	1.143	.760	1.719
Social Support	.000	.207	.000	1	.999	1.000	.666	1.501
Social Control	.316	.223	2.021	1	.155	1.372	.887	2.122

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I3

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Technical Violations for Hispanic Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.976	.839	1.354	1	.245	2.654	.513	13.740
Attitudes towards Authority	1.420	.695	4.174	1	**.041**	4.137	1.059	16.158
Impulse Control	.701	.752	.870	1	.351	2.016	.462	8.801
Problem Solving	.889	.696	1.632	1	.201	2.432	.622	9.511
Sense of Entitlement	.091	.561	.026	1	.871	1.095	.365	3.290
Attachment with Others	-.382	.676	.319	1	.572	.682	.181	2.569
Substance Abuse	-.361	.609	.351	1	.554	.697	.211	2.301
Anger	-.768	.648	1.405	1	.236	.464	.130	1.652
Access to Victims	-1.291	.593	4.742	1	**.029**	.275	.086	.879
Negative Mood	.002	.486	.000	1	.996	1.002	.387	2.598
Employment	.385	.478	.650	1	.420	1.470	.576	3.750
Interpersonal Relationships	1.281	.608	4.441	1	**.035**	3.600	1.094	11.852
Living Situation	-.503	.473	1.131	1	.288	.605	.239	1.529
Response to advice	1.387	.981	1.998	1	.158	4.003	.585	27.402
Prosocial Identity	-1.394	.851	2.682	1	.102	.248	.047	1.316
High Expectations	.773	.715	1.168	1	.280	2.166	.533	8.797
Costs/Benefits	-.727	.823	.780	1	.377	.483	.096	2.427
Social Support	.974	.762	1.634	1	.201	2.648	.595	11.785
Social Control	-.942	.876	1.155	1	.282	.390	.070	2.172

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I4

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Technical Violations for White Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.141	.087	2.639	1	.104	1.152	.971	1.365
Attitudes towards Authority	-.067	.088	.576	1	.448	.935	.787	1.111
Impulse Control	.074	.097	.580	1	.446	1.076	.891	1.300
Problem Solving	.198	.096	4.245	1	**.039**	1.219	1.010	1.472
Sense of Entitlement	.042	.083	.257	1	.612	1.043	.886	1.227
Attachment with Others	-.119	.083	2.061	1	.151	.888	.755	1.044
Substance Abuse	.186	.067	7.698	1	**.006**	1.204	1.056	1.373
Anger	.075	.076	.985	1	.321	1.078	.929	1.251
Access to Victims	-.040	.071	.328	1	.567	.960	.836	1.103
Negative Mood	.015	.069	.049	1	.825	1.015	.887	1.163
Employment	.133	.060	4.923	1	**.027**	1.143	1.016	1.286
Interpersonal Relationships	.110	.073	2.303	1	.129	1.116	.968	1.287
Living Situation	.024	.070	.119	1	.730	1.025	.893	1.176
Response to advice	-.109	.098	1.240	1	.265	.897	.740	1.086
Prosocial Identity	-.203	.102	3.968	1	**.046**	.816	.668	.997
High Expectations	.174	.097	3.245	1	.072	1.190	.985	1.438
Costs/Benefits	.041	.098	.173	1	.678	1.042	.859	1.263
Social Support	.022	.095	.056	1	.814	1.023	.848	1.233
Social Control	-.082	.103	.633	1	.426	.921	.753	1.127

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I5

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Technical Violations for Black Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	-.055	.199	.076	1	.783	.947	.640	1.399
Attitudes towards Authority	.410	.189	4.704	1	**.030**	1.507	1.040	2.182
Impulse Control	-.129	.209	.383	1	.536	.879	.584	1.323
Problem Solving	.263	.209	1.582	1	.209	1.301	.863	1.962
Sense of Entitlement	.091	.175	.270	1	.603	1.095	.778	1.542
Attachment with Others	-.280	.185	2.286	1	.131	.756	.525	1.087
Substance Abuse	.192	.138	1.937	1	.164	1.212	.925	1.589
Anger	-.132	.160	.676	1	.411	.876	.640	1.200
Access to Victims	.082	.161	.260	1	.610	1.086	.792	1.489
Negative Mood	-.143	.154	.863	1	.353	.867	.641	1.172
Employment	.096	.134	.519	1	.471	1.101	.847	1.431
Interpersonal Relationships	.117	.165	.502	1	.479	1.124	.813	1.554
Living Situation	.036	.158	.053	1	.819	1.037	.761	1.413
Response to advice	.198	.211	.882	1	.348	1.219	.806	1.845
Prosocial Identity	-.058	.236	.060	1	.807	.944	.594	1.499
High Expectations	.121	.220	.302	1	.583	1.128	.734	1.735
Costs/Benefits	-.123	.230	.287	1	.592	.884	.563	1.389
Social Support	.012	.238	.003	1	.958	1.013	.635	1.614
Social Control	-.061	.244	.063	1	.802	.940	.583	1.518

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I6

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Technical Violations for Hispanic Women

							95% CI	
DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	Lower	Upper
Peer Associations	.576	.423	1.849	1	.174	1.778	.776	4.078
Attitudes towards Authority	.602	.438	1.889	1	.169	1.826	.774	4.307
Impulse Control	.362	.496	.532	1	.466	1.436	.543	3.800
Problem Solving	-.338	.492	.470	1	.493	.714	.272	1.872
Sense of Entitlement	.031	.516	.004	1	.951	1.032	.375	2.836
Attachment with Others	.211	.464	.206	1	.650	1.234	.497	3.066
Substance Abuse	.431	.400	1.157	1	.282	1.538	.702	3.371
Anger	.345	.394	.764	1	.382	1.411	.652	3.056
Access to Victims	-.003	.383	.000	1	.994	.997	.471	2.112
Negative Mood	-.244	.382	.406	1	.524	.784	.371	1.658
Employment	-.290	.343	.716	1	.398	.748	.382	1.466
Interpersonal Relationships	.088	.414	.045	1	.832	1.092	.485	2.456
Living Situation	-.087	.407	.046	1	.830	.917	.413	2.035
Response to advice	.017	.536	.001	1	.974	1.017	.356	2.910
Prosocial Identity	.918	.650	1.992	1	.158	2.503	.700	8.952
High Expectations	-.234	.488	.230	1	.632	.791	.304	2.059
Costs/Benefits	.244	.484	.255	1	.614	1.277	.495	3.294
Social Support	.108	.481	.051	1	.822	1.114	.434	2.860
Social Control	-.310	.613	.256	1	.613	.733	.221	2.436

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I7

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting New Offences for White Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.133	.117	1.291	1	.256	1.142	.908	1.436
Attitudes towards Authority	.172	.109	2.461	1	.117	1.187	.958	1.471
Impulse Control	.166	.131	1.604	1	.205	1.180	.913	1.525
Problem Solving	.133	.129	1.065	1	.302	1.142	.887	1.470
Sense of Entitlement	.024	.103	.054	1	.816	1.024	.838	1.252
Attachment with Others	.005	.114	.002	1	.964	1.005	.804	1.256
Substance Abuse	-.059	.095	.388	1	.533	.943	.783	1.135
Anger	-.113	.098	1.321	1	.250	.893	.736	1.083
Access to Victims	-.137	.095	2.085	1	.149	.872	.723	1.050
Negative Mood	.146	.091	2.548	1	.110	1.157	.967	1.384
Employment	.106	.082	1.675	1	.196	1.112	.947	1.307
Interpersonal Relationships	.089	.095	.881	1	.348	1.093	.908	1.317
Living Situation	-.122	.093	1.718	1	.190	.885	.738	1.062
Response to advice	-.015	.131	.014	1	.907	.985	.762	1.272
Prosocial Identity	.009	.142	.004	1	.952	1.009	.763	1.333
High Expectations	-.221	.130	2.900	1	.089	.802	.621	1.034
Costs/Benefits	.152	.133	1.306	1	.253	1.165	.897	1.513
Social Support	.121	.121	1.000	1	.317	1.128	.891	1.429
Social Control	-.007	.137	.003	1	.958	.993	.760	1.298

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I8

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting New Offences for Black Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	-.067	.209	.102	1	.750	.935	.620	1.410
Attitudes towards Authority	-.012	.207	.004	1	.952	.988	.658	1.482
Impulse Control	-.003	.235	.000	1	.990	.997	.629	1.580
Problem Solving	.107	.243	.194	1	.659	1.113	.692	1.790
Sense of Entitlement	.231	.202	1.312	1	.252	1.260	.849	1.871
Attachment with Others	.054	.212	.065	1	.798	1.056	.697	1.599
Substance Abuse	-.042	.166	.064	1	.800	.959	.693	1.327
Anger	-.132	.191	.478	1	.489	.876	.602	1.275
Access to Victims	-.236	.172	1.886	1	.170	.790	.564	1.106
Negative Mood	-.084	.184	.211	1	.646	.919	.641	1.317
Employment	.103	.156	.441	1	.507	1.109	.817	1.505
Interpersonal Relationships	.317	.186	2.901	1	.089	1.372	.953	1.976
Living Situation	.045	.171	.069	1	.793	1.046	.748	1.462
Response to advice	-.295	.244	1.468	1	.226	.744	.461	1.200
Prosocial Identity	-.241	.273	.780	1	.377	.786	.460	1.341
High Expectations	.028	.241	.014	1	.907	1.029	.642	1.648
Costs/Benefits	.026	.251	.011	1	.917	1.027	.628	1.679
Social Support	.097	.255	.144	1	.704	1.102	.668	1.816
Social Control	.174	.270	.415	1	.520	1.190	.701	2.018

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I9

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting New Offences for Hispanic Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	-.273	1.734	.025	1	.875	.761	.025	22.756
Attitudes towards Authority	.762	1.077	.500	1	.480	2.142	.259	17.695
Impulse Control	.281	1.418	.039	1	.843	1.324	.082	21.336
Problem Solving	.570	1.258	.205	1	.651	1.767	.150	20.794
Sense of Entitlement	-2.861	1.740	2.704	1	.100	.057	.002	1.731
Attachment with Others	-.053	1.249	.002	1	.966	.948	.082	10.965
Substance Abuse	-2.683	1.290	4.325	1	**.038**	.068	.005	.857
Anger	-.544	1.169	.216	1	.642	.580	.059	5.741
Access to Victims	.676	1.421	.226	1	.634	1.966	.121	31.873
Negative Mood	-2.420	1.892	1.635	1	.201	.089	.002	3.628
Employment	-.341	.943	.131	1	.718	.711	.112	4.514
Interpersonal Relationships	1.490	1.125	1.757	1	.185	4.439	.490	40.220
Living Situation	1.372	.984	1.942	1	.163	3.942	.573	27.140
Response to advice	-.496	1.615	.094	1	.759	.609	.026	14.439
Prosocial Identity	1.417	1.712	.685	1	.408	4.125	.144	118.282
High Expectations	-2.973	2.387	1.552	1	.213	.051	.000	5.498
Costs/Benefits	-1.881	1.653	1.296	1	.255	.152	.006	3.889
Social Support	1.124	1.211	.862	1	.353	3.078	.287	33.012
Social Control	-.076	1.735	.002	1	.965	.927	.031	27.807

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I10

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting New Offences for White Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.307	.135	5.154	1	**.023**	1.359	1.043	1.772
Attitudes towards Authority	-.140	.128	1.192	1	.275	.869	.676	1.118
Impulse Control	.039	.149	.067	1	.796	1.039	.776	1.392
Problem Solving	.093	.146	.405	1	.525	1.098	.824	1.462
Sense of Entitlement	.067	.120	.309	1	.578	1.069	.845	1.351
Attachment with Others	-.079	.121	.431	1	.511	.924	.729	1.170
Substance Abuse	-.067	.105	.410	1	.522	.935	.762	1.148
Anger	.219	.109	4.058	1	**.044**	1.245	1.006	1.541
Access to Victims	-.141	.105	1.812	1	.178	.869	.708	1.066
Negative Mood	-.007	.103	.005	1	.944	.993	.811	1.216
Employment	.024	.091	.070	1	.791	1.024	.858	1.224
Interpersonal Relationships	-.276	.107	6.699	1	**.010**	.759	.615	.935
Living Situation	-.061	.104	.346	1	.557	.941	.768	1.153
Response to advice	-.194	.146	1.772	1	.183	.824	.619	1.096
Prosocial Identity	-.102	.151	.453	1	.501	.903	.672	1.214
High Expectations	.239	.141	2.861	1	.091	1.270	.963	1.676
Costs/Benefits	.075	.146	.267	1	.605	1.078	.811	1.434
Social Support	-.176	.141	1.555	1	.212	.839	.636	1.106
Social Control	-.141	.153	.847	1	.357	.868	.643	1.173

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I11

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting New Offences for Black Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	-.081	.284	.081	1	.776	.922	.528	1.610
Attitudes towards Authority	.067	.263	.066	1	.797	1.070	.639	1.790
Impulse Control	.089	.317	.079	1	.778	1.093	.587	2.037
Problem Solving	-.051	.299	.029	1	.865	.950	.528	1.709
Sense of Entitlement	.167	.240	.488	1	.485	1.182	.739	1.891
Attachment with Others	.380	.265	2.056	1	.152	1.462	.870	2.456
Substance Abuse	-.033	.195	.028	1	.866	.968	.661	1.417
Anger	.112	.239	.219	1	.640	1.118	.700	1.784
Access to Victims	-.041	.224	.033	1	.856	.960	.619	1.489
Negative Mood	.381	.224	2.893	1	.089	1.464	.944	2.271
Employment	.004	.193	.000	1	.984	1.004	.687	1.467
Interpersonal Relationships	.159	.237	.455	1	.500	1.173	.738	1.865
Living Situation	-.065	.222	.085	1	.771	.937	.607	1.448
Response to advice	-.033	.301	.012	1	.912	.967	.536	1.746
Prosocial Identity	.378	.336	1.265	1	.261	1.459	.755	2.820
High Expectations	.063	.310	.041	1	.840	1.065	.579	1.956
Costs/Benefits	.109	.335	.105	1	.746	1.115	.578	2.151
Social Support	.280	.333	.710	1	.399	1.324	.690	2.541
Social Control	-.584	.346	2.845	1	.092	.558	.283	1.099

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I12

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Any Return for White Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.109	.081	1.822	1	.177	1.116	.952	1.308
Attitudes towards Authority	.023	.079	.082	1	.774	1.023	.876	1.195
Impulse Control	.118	.089	1.729	1	.188	1.125	.944	1.340
Problem Solving	.159	.090	3.138	1	.077	1.173	.983	1.399
Sense of Entitlement	-.014	.075	.033	1	.855	.986	.851	1.143
Attachment with Others	-.023	.083	.078	1	.780	.977	.831	1.149
Substance Abuse	.193	.065	8.795	1	**.003***	1.212	1.067	1.377
Anger	.010	.070	.020	1	.887	1.010	.880	1.159
Access to Victims	-.090	.068	1.767	1	.184	.914	.801	1.044
Negative Mood	.091	.066	1.907	1	.167	1.095	.963	1.246
Employment	.103	.058	3.089	1	.079	1.108	.988	1.243
Interpersonal Relationships	.337	.069	24.195	1	**< .001***	1.401	1.225	1.603
Living Situation	-.184	.068	7.340	1	**.007**	.832	.728	.950
Response to advice	-.057	.093	.380	1	.537	.944	.787	1.133
Prosocial Identity	-.068	.102	.444	1	.505	.935	.766	1.141
High Expectations	-.220	.093	5.655	1	**.017**	.802	.669	.962
Costs/Benefits	.055	.095	.338	1	.561	1.057	.877	1.274
Social Support	.082	.088	.862	1	.353	1.086	.913	1.291
Social Control	-.001	.098	.000	1	.992	.999	.825	1.210

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I13

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Any Return for Black Men

							95% CI	
DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	Lower	Upper
Peer Associations	.194	.173	1.258	1	.262	1.214	.865	1.703
Attitudes towards Authority	-.120	.173	.481	1	.488	.887	.632	1.245
Impulse Control	-.005	.196	.001	1	.980	.995	.678	1.460
Problem Solving	.080	.202	.156	1	.693	1.083	.729	1.609
Sense of Entitlement	.409	.177	5.330	1	**.021**	1.505	1.064	2.129
Attachment with Others	-.231	.180	1.648	1	.199	.793	.557	1.130
Substance Abuse	.332	.137	5.909	1	**.015**	1.394	1.066	1.821
Anger	-.225	.159	1.990	1	.158	.799	.584	1.092
Access to Victims	-.123	.145	.724	1	.395	.884	.666	1.174
Negative Mood	-.037	.152	.059	1	.808	.964	.715	1.299
Employment	.002	.128	.000	1	.987	1.002	.779	1.288
Interpersonal Relationships	.681	.164	17.288	1	**< .001***	1.977	1.434	2.725
Living Situation	-.387	.142	7.413	1	**.006**	.679	.514	.897
Response to advice	-.216	.203	1.131	1	.288	.805	.541	1.200
Prosocial Identity	.157	.222	.499	1	.480	1.170	.757	1.809
High Expectations	-.540	.207	6.835	1	**.009**	.583	.389	.874
Costs/Benefits	.169	.213	.634	1	.426	1.185	.781	1.798
Social Support	.109	.210	.268	1	.604	1.115	.739	1.681
Social Control	.352	.226	2.425	1	.119	1.421	.913	2.212

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I14

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Any Return for Hispanic Men

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.961	.836	1.322	1	.250	2.614	.508	13.444
Attitudes towards Authority	1.606	.720	4.973	1	**.026**	4.982	1.215	20.437
Impulse Control	.417	.773	.291	1	.590	1.517	.334	6.900
Problem Solving	.819	.704	1.354	1	.245	2.268	.571	9.008
Sense of Entitlement	.186	.564	.108	1	.742	1.204	.399	3.636
Attachment with Others	-.587	.700	.705	1	.401	.556	.141	2.190
Substance Abuse	-.208	.604	.118	1	.731	.812	.249	2.655
Anger	-.711	.646	1.209	1	.272	.491	.138	1.744
Access to Victims	-1.197	.600	3.984	1	**.046**	.302	.093	.979
Negative Mood	-.100	.495	.041	1	.839	.904	.343	2.385
Employment	.284	.485	.342	1	.559	1.328	.513	3.439
Interpersonal Relationships	1.333	.605	4.862	1	**.027**	3.793	1.160	12.405
Living Situation	-.351	.480	.536	1	.464	.704	.275	1.802
Response to advice	1.399	.979	2.043	1	.153	4.050	.595	27.569
Prosocial Identity	-1.285	.855	2.259	1	.133	.277	.052	1.478
High Expectations	.853	.723	1.392	1	.238	2.346	.569	9.673
Costs/Benefits	-1.000	.846	1.395	1	.237	.368	.070	1.933
Social Support	1.077	.762	1.999	1	.157	2.936	.660	13.066
Social Control	-.995	.871	1.305	1	.253	.370	.067	2.038

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I15

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Any Return for White Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.159	.088	3.292	1	.070	1.173	.987	1.393
Attitudes towards Authority	-.025	.089	.079	1	.779	.975	.818	1.162
Impulse Control	.098	.098	1.005	1	.316	1.103	.911	1.335
Problem Solving	.184	.097	3.559	1	.059	1.202	.993	1.454
Sense of Entitlement	.051	.084	.367	1	.544	1.052	.892	1.242
Attachment with Others	-.152	.084	3.292	1	.070	.859	.729	1.012
Substance Abuse	.170	.068	6.291	1	**.012**	1.185	1.038	1.353
Anger	.070	.077	.824	1	.364	1.072	.922	1.247
Access to Victims	-.050	.072	.480	1	.489	.952	.827	1.095
Negative Mood	.017	.070	.059	1	.808	1.017	.887	1.167
Employment	.140	.061	5.319	1	**.021**	1.151	1.021	1.297
Interpersonal Relationships	.083	.074	1.283	1	.257	1.087	.941	1.256
Living Situation	.012	.071	.030	1	.863	1.012	.880	1.164
Response to advice	-.104	.099	1.091	1	.296	.901	.742	1.095
Prosocial Identity	-.202	.103	3.808	1	**.050**	.817	.667	1.001
High Expectations	.196	.098	4.017	1	**.045**	1.217	1.004	1.475
Costs/Benefits	.014	.100	.020	1	.886	1.014	.834	1.233
Social Support	.013	.097	.018	1	.894	1.013	.838	1.224
Social Control	-.102	.104	.957	1	.328	.903	.736	1.108

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I16

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Any Return for Black Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	-.081	.204	.156	1	.693	.922	.618	1.376
Attitudes towards Authority	.427	.194	4.851	1	**.028**	1.533	1.048	2.242
Impulse Control	-.137	.214	.413	1	.521	.872	.573	1.325
Problem Solving	.325	.214	2.305	1	.129	1.385	.910	2.108
Sense of Entitlement	.061	.179	.116	1	.734	1.063	.749	1.508
Attachment with Others	-.326	.190	2.957	1	.086	.722	.497	1.047
Substance Abuse	.204	.142	2.076	1	.150	1.226	.929	1.618
Anger	-.093	.164	.318	1	.573	.911	.660	1.258
Access to Victims	.049	.165	.090	1	.765	1.051	.760	1.452
Negative Mood	-.078	.157	.243	1	.622	.925	.680	1.259
Employment	.086	.137	.398	1	.528	1.090	.834	1.426
Interpersonal Relationships	.202	.169	1.420	1	.233	1.224	.878	1.705
Living Situation	.011	.162	.005	1	.945	1.011	.736	1.390
Response to advice	.223	.216	1.064	1	.302	1.250	.818	1.909
Prosocial Identity	-.141	.242	.339	1	.560	.869	.540	1.396
High Expectations	.214	.226	.898	1	.343	1.239	.795	1.930
Costs/Benefits	-.083	.236	.123	1	.726	.920	.579	1.462
Social Support	.059	.244	.059	1	.809	1.061	.657	1.712
Social Control	-.115	.252	.210	1	.647	.891	.544	1.459

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.

Table I17

Logistic Regression Analysis Coefficients for Dynamic Risk Assessment for Offender Re-Entry Items – Predicting Any Return for Hispanic Women

DRAOR Item	*B*	S.E.	Wald	*df*	Sig.	Odds Ratio	95% CI Lower	Upper
Peer Associations	.619	.448	1.906	1	.167	1.857	.771	4.471
Attitudes towards Authority	.715	.466	2.351	1	.125	2.044	.820	5.098
Impulse Control	.689	.515	1.785	1	.182	1.991	.725	5.469
Problem Solving	-.135	.506	.071	1	.790	.874	.324	2.357
Sense of Entitlement	-.089	.539	.027	1	.869	.915	.318	2.631
Attachment with Others	.043	.480	.008	1	.929	1.044	.407	2.675
Substance Abuse	.487	.423	1.325	1	.250	1.627	.710	3.729
Anger	.022	.427	.003	1	.958	1.023	.443	2.362
Access to Victims	.217	.399	.297	1	.586	1.243	.568	2.717
Negative Mood	-.492	.410	1.442	1	.230	.611	.274	1.365
Employment	-.565	.371	2.318	1	.128	.569	.275	1.176
Interpersonal Relationships	.073	.428	.029	1	.864	1.076	.465	2.489
Living Situation	-.017	.421	.002	1	.967	.983	.431	2.243
Response to advice	-.135	.552	.060	1	.807	.874	.296	2.578
Prosocial Identity	1.072	.686	2.440	1	.118	2.920	.761	11.204
High Expectations	-.086	.511	.029	1	.866	.917	.337	2.497
Costs/Benefits	.485	.513	.896	1	.344	1.625	.595	4.440
Social Support	.158	.511	.095	1	.757	1.171	.430	3.189
Social Control	-.856	.674	1.615	1	.204	.425	.113	1.591

Note. **Bold** denotes significance at the alpha = .05 level; **Bold*** denotes significance following correction to *p* values to account for multiple corrections.