Predicting Women's Recidivism using the Dynamic Risk Assessment for Offender Re-entry: Preliminary Evidence of Predictive Validity with Community-Sentenced Women using a 'Gender-neutral' Risk Measure

Jessica M. Scanlan

Julia A. Yesberg

Clare-Ann Fortune

Devon L. L. Polaschek

Victoria University of Wellington

Authors' Note

We would like to acknowledge, with gratitude, the permission granted by Dr Ralph Serin, the author of the DRAOR, to collate, analyze and report on the New Zealand DRAOR data in this manuscript. We would also like to acknowledge the invaluable support of the New Zealand Department of Corrections.

Abstract

Although men and women share risk factors for offending, some scholars suggest these factors operate differently across gender and that women-specific risk factors are neglected in existing 'gender-neutral' risk assessment tools. This paper explored the predictive validity of one gender-neutral risk assessment tool—the Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007; Serin, Mailloux, & Wilson, 2012)—with matched samples of women and men serving community supervision sentences. Total DRAOR scores had comparative predictive validity across gender. For women and men, the DRAOR predicted reconviction over a static risk measure. The findings support the general premise of gender neutrality, but do not necessarily suggest the DRAOR, or gender-neutral tools more broadly, are the best tools for use with women.

Keywords: risk assessment, DRAOR, gender, recidivism, women

Introduction

Over the past four decades there has been a shift in the methods used to produce criminal risk assessments, but assessment for women has lagged well behind developments in risk assessment for men. Today, actuarial tools are most commonly used to generate assessments. These tools evaluate empirically and theoretically derived correlates of criminal behavior to produce a numerical score which places an individual in a category based on his or her risk of future offending. Risk assessment is central to the effective management of people under the aegis of criminal justice (CJ) systems, and informs a range of decisions, such as their parole and/or treatment eligibility, treatment targets, and degree of punitive sanctions imposed: people at low risk of criminal reconviction are more likely to be given community sentences for the same offense than their higher risk cohorts, for offenses with sentencing discretion. Arriving at an accurate risk assessment is a challenging but fundamentally important task, and ensuring tools are valid and reliable predictors for all populations—including the growing population of women—is of increasing significance. Also of increasing importance is the value of instruments that are effective with people being managed in the community, particularly those that have the potential to inform management. This study reports on a dynamic assessment instrument used by community correctional staff in the management of women and men serving community supervision (i.e., probation-type) sentences in New Zealand.

In New Zealand, the number of women in prisons is growing at a faster rate than for men (Department of Corrections, 2017a). The number of incarcerated women almost tripled between 2002 and 2017 (275 and 800 respectively; Department of Corrections, 2017b; Statistics New Zealand, 2012). Similar trends are observed internationally. In the US, the number of female sentenced prisoners under the jurisdiction of either state or federal correctional authorities increased 0.7 percent between 2015 and 2016 while the number of male sentenced prisoners over the same period saw a 1.3 percent decrease (Carson, 2018). Although trends show an increase in the number of women entering CJ systems, globally, men still comprise the majority of those serving sentences. This 'gender gap' in offending is observed over time, culture, and country (Steffensmeier & Allan, 1996) and has prioritized research efforts towards men: the result of which has been a proliferation of research concerning the assessment of risk and the development of risk assessment tools that almost exclusively focuses on men. The relative neglect of women in the published risk assessment literature has led scholars to question the generalizability of risk assessment tools developed with men to the prediction of women's criminal risk (Van Voorhis, 2012). This paper explores the generalizability of such a risk assessment tool across gender, employing a matched-sample methodology.

Criminal Risk Assessment

Since the unstructured clinical judgement approach of the 1960s that yielded predictions often no better than chance (Grove, Zald, Lebow, Snitz, & Nelson, 2000), the field of risk assessment has advanced greatly. Risk scales, scored actuarially, are now the norm. These scales typically stipulate strict sets of instructions for the evaluation and formulation of assessments using empirically derived risk factors. Risk factors are observable or measurable proxies which account for variance in the unobservable construct of future criminal behavior.

Scales built from static factors typically rely on historic markers that change slowly (e.g., age) or are largely unchangeable by human effort (e.g., age of first conviction). More recent measures incorporate personal and environmentally-based dynamic risk factors, including antisocial peers, impulsivity, and antisocial cognitions (Bonta & Andrews, 2017). Measures that contain them can be scored so as to capture the actual circumstances of the assessed person, and have uses well beyond recidivism prediction (Yesberg & Polaschek, 2015). They are useful for documenting treatment-related change (e.g., Olver, Wong, Nicholaichuk & Gordon, 2007), and can capture unique variance in recidivism after controlling for static risk (Dickson, Polaschek, & Casey, 2013; Olver, & Jung, 2017). Present day risk assessment tools are largely dynamic in nature, but many continue to include static risk factors because of their strong empirical relationship with recidivism.

Even more recently, interest has grown in factors that decrease a person's likelihood of engaging in antisocial acts: these factors are referred to as protective factors. There is considerable debate and insufficient research on the exact nature of protective factors (Polaschek, 2017; Serin, Chadwick, & Lloyd, 2016; Walker, Bowen, & Brown, 2013) and how they function (de Vries Robbe, 2014; Fortune & Ward, 2017; Lösel & Farrington, 2012; Jones, Brown, Robinson, & Frey, 2015; Spice, Viljoen, Lazman, Scalora, & Ullman, 2012; Ullrich & Coid, 2011). One common definition is that they are strengths or resources, internal or external to a person, that mitigate against engaging in criminal behavior. Protective factors have been incorporated into dynamic risk assessment tools with the intention of providing a more balanced appraisal of risk. Empirical studies with one such scale have supported the value of including protective factors in structured risk assessments (the SAPROF; de Vries Robbé, de Vogel, & de Spa, 2011; de Vries Robbé, de Vogel, & Douglas, 2013), although the question of whether they are distinct from risk factors conceptually remains largely unresolved to date.

Two Perspectives on Risk Assessment for Women

Most actuarial risk scales used in contemporary correctional services were originally developed on samples of men and later applied to women (Blanchette & Taylor, 2007). The principal assumption underpinning this practice is that risk factors are the same for men and women and that tools developed on men are valid for use with women. However, there is a noticeable lack of research exploring the veracity of these assumptions (Zakaria, Allenby, Derkzen, & Jones, 2013). A review of the literature reveals two distinct schools of thought regarding the assessment of women's risk: the gender-neutral and gender-responsive perspectives.

Gender-Neutral Perspective. The gender-neutral perspective is theoretically grounded in Bonta and Andrews' General Personality and Cognitive Social Learning (GPCSL) model of criminal behavior (Bonta & Andrews, 2017). The GPCSL model recognizes the commission of criminal behavior as the result of an interaction between people and their immediate environments. According to this perspective, a person's gender is a distal influence, residing in the broader socio-political context, and consequentially has minimal impact on criminality after the consideration of more immediate factors (Rettinger & Andrews, 2010). These immediate factors are known as the 'Central Eight' risk factors, which are primarily dynamic in nature and have been identified as the strongest among such predictors of criminal behavior (Bonta & Andrews, 2017). Andrews and Bonta previously differentiated between the Big Four (history of antisocial behavior, antisocial personality pattern, antisocial cognitions, antisocial associates) and the Moderate Four (family/marital, school/work, leisure/recreation, and substance abuse) risk factors but have removed this distinction in the most recent edition (Bonta & Andrews, 2017).

Empirical support for the gender-neutral perspective comes from research showing scales first developed for use with men predict recidivism of women; for example, the widely used Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 2010), developed in accordance with the GPCSL. In a large-scale meta-analysis of the Level of Service scales, Olver, Stockdale and Wormith (2014) found approximately equivalent predictive accuracy with men and women. An earlier meta-analysis of 27 studies of the LSI-R found it to have acceptable predictive validity with women (r = .35, 95% CI [.34. .36]; Smith, Cullen, & Latessa, 2009). Similarly, a recent meta-analysis with young people found comparable predictive validity for the Youth Level of Service/Case Management Inventory (YLS/CMI) for boys and girls (*r* = .28 and .25, respectively for general recidivism; Pusch & Holtfreter, 2018). Geraghty and Woodhams (2015) also confirmed the predictive validity of the LS measures with women in a systematic review, but found a number of other risk measures developed for use with men did not reach acceptable levels of accuracy with women. Finally, in a small scale study of young men and women, the YLS/CMI predicted recidivism equally well for both; however, the matching of services to criminogenic needs was only associated with reduced recidivism for males (Vitopoulos, Peterson-Badali, & Skilling, 2012).

Critiquing Gender-Neutrality from a Gender-Responsive Perspective. Despite empirical and theoretical support for a gender-neutral perspective, women serving sentences in the criminal justice system are considerably different from their male counterparts in terms of the type, severity and frequency of criminal behavior. Women participate in crime at much lower rates than men, are less violent, and may be more likely to engage in crime involving alcohol, other drugs, or property (Bloom, Owen, & Covington, 2003). According to the gender-responsive perspective women are thus a distinct offending population, with gender-specific experiences that led to their involvement in antisocial behavior, and as a result, present with risk factors distinct from those of men. Proponents of the gender-responsive perspective question the validity of using contemporary actuarial assessment scales with women because they neglect or fail to accurately weight factors considered pertinent to female criminality (Blanchette & Brown, 2006; Reisig, Holtfreter, & Morash, 2006). Proposed gendered risk factors variously include history of victimization, relationship dysfunction, mental health problems, substance abuse, self-efficacy and parental stress (Wright, Salisbury, & Van Voorhis, 2007; Van Voorhis, Wright, Salisbury, & Bauman, 2010).

There is mixed empirical support for women-specific risk factors, providing some basis for challenging gender neutrality arguments. For example, some studies show that although predictive overall, some measures may achieve accuracy based on different component risks for men and women. This research base has found some differences in which factors within existing measures predict recidivism for men and women (Baglivio & Jackowski, 2013; Nicholls, Ogloff, & Douglas, 2004). For example, LSI-R risk factors related to accommodation, education, work and relationships with friends have been found to be more strongly correlated with general recidivism for men relative to women; in contrast, difficulties with emotional wellbeing was more strongly related to women's recidivism than men's (Olver et al., 2014; van der Knaap, Alberda, Oosterveld, & Born, 2012).

Research also shows a mixed picture with regard to the predictiveness and incremental predictive ability of women-specific factors. Van Voorhis and colleagues (2010) examined the incremental predictive validity of different combinations of gender-responsive supplements over and above the predictive ability of the LSI-R in prison, parole, and probation samples of women. They found that in six of the eight samples studied, gender-responsive scales accounted for unique predictive variance above the gender-neutral models. But there was considerable variation in which factors added to accuracy across studies. Similarly, in an important study that compared men and women on a variety of risk factors derived from the Van Voorhis' supplements, Bell (2014) found that of eight gender-neutral risk items, five were significant predictors for men, and three (including two that were predictive for men) were predictive for women. Four of nine women-specific risk factors were predictive for women (and three for men, but two of these were predictive in the opposite direction to what was theorized).

Reisig and colleagues (2006) coded women's criminal behavior on Daly's (1992) pathways framework and subsequently examined the predictive validity of the LSI-R.[1] They found the LSI-R predicted recidivism for women who followed the 'economically motivated' pathway, but failed to accurately predict recidivism for women on the gendered pathways.

This pattern raises the possibility that a hybrid perspective (Van Voorhis, 2012), based on a better understanding of heterogeneity in women's pathways to crime, holds promise. In this way of thinking, some risk factors for men will predict equally well for (some) women, but within the population of women who commit crimes, there will be others who do not fit the gender-neutral (i.e., male) mold, and women may also differ within themselves in respect of which risk factors are most important in prediction and management. So, following from Daly's (1992) seminal study, more recent pathways research on women confirms considerable diversity in the routes by which women have come into the criminal justice system (see Belknap & Holsinger, 2006; Wattanaporn & Holtfreter, 2014). For example, using a person-centered statistical approach, Brennan, Breitenbach, Dieterich, Salisbury and Van Voorhis (2012) found eight pathways, several of which appeared to be more typical of women than men; though no links were made between these pathways and recidivism prediction.

There is also evidence women are more likely to have higher "welfare" and current safety needs (i.e., difficulties that arguably should be attended to because they affect every-day wellbeing and adaptive functioning—such as lack of housing, and ongoing victimization—but may not necessarily have a direct relationship to recidivism). For example, depression, housing safety and lack of self-efficacy were all found to be higher in women than men in Bell's (2014) study, but none predicted recidivism. Relatedly, interventions focusing on recovery from trauma tend to focus on the wellbeing benefits, not the potential for recidivism reduction (Kubiak, Kim, Fedock, & Bybee, 2012; Roe-Sepowitz, Bedard, Pate, & Hedberg, 2014).

Background to the Present Study

Although this is still very much an emerging area of empirical research, existing studies suggest there is potential for women-specific risk factors to complement those traditionally found to be predictive for men. Research that clusters risk factors, or arranges them retrospectively into developmental pathways, suggests considerable diversity within the population of women with offending histories, perhaps more than is typical for men. But these studies still find some women whose offending patterns are a good fit to pathways originally developed for men. This observation raises the possibility that the population under study will be important in determining findings for individual studies, as will the ability of different risk instruments to accommodate these variations. Based on the few meta-analyses available, the LSI measures may be more capable of accommodating women's heterogeneity in offending-related risks than some other scales.

Regardless, for correctional systems, being able to use the same instrument is efficient and practical for a number of reasons, suggesting that determining whether they do a "good enough" job is important in a pragmatic sense. For example, developing separate instruments is a costly business, as is training staff to standard in their use. Given ongoing debate regarding the applicability of existing gender-neutral tools to women, this paper assessed the predictive validity of a gender-neutral assessment tool—the Dynamic Risk Assessment for Offender Re-entry (DRAOR, see measures section; Serin, 2007; Serin, Mailloux, & Wilson, 2012)—with a community-sentenced sample of women and a matched sample of men. Since its development, the DRAOR's convergent, predictive, and incremental predictive validity has been investigated in a number of studies using representative samples of New Zealand adult men (Hanby, 2013; Tamatea & Wilson, 2009; Yesberg & Polaschek, 2015). Additionally, the DRAOR has been evaluated in a mixed sample of people on sentence in the United States (Chadwick, 2014). These studies have shown the DRAOR to be useful at distinguishing men who are likely to be reconvicted from men who are not (e.g., AUC = .62 for new convictions excluding breaches; Yesberg & Polaschek, 2015).

To date only one study has directly examined the predictive validity of the DRAOR on matched samples of women and men (Yesberg, Scanlan, Hanby, Serin, & Polaschek, 2015).

Yesberg and colleagues (2015) examined the predictive validity of the DRAOR with a sample of 133 women and a matched sample of 133 men serving parole sentences in New Zealand. The authors found the DRAOR was able to predict recidivism for both samples. For women, the Acute subscale made a significant unique contribution to the multivariate subscale model; however, the same model for men did not predict recidivism. Finally, the paper found the DRAOR made an additional significant contribution above a static risk assessment tool (the RoC*RoI, see measures section) when predicting women's recidivism but not men's.

This study builds on Yesberg and colleagues' (2015) parole study, employing a new and larger sample of women and men who were serving community-based supervision sentences (i.e., not parole). This study also used a more statistically rigorous methodology for matching the two samples. The following three research questions were addressed: (1) Does the DRAOR predict reconviction with equal accuracy for a matched sample of community-sentenced women and men? (2) Are there differences in which subscales underpin predictive validity for women vs. men? And (3) Does the DRAOR provide incremental predictive validity above a static-factor based risk estimate for women and for men?

Method

Data and Sample Preparation

The data set was drawn from a New Zealand Department of Corrections database of all people serving a community supervision sentence that began after 1 January 2011 and ended prior to 31 December 2013. The original data set included 1,100 randomly selected adults on community sentences: 550 women and 550 men. After removing cases with insufficient risk assessment data (e.g., missing RoC*RoI score, fewer than 5 DRAOR scores over sentence, first DRAOR score more than 6 weeks into sentence, more than two months between any two DRAOR scores)[2], 336 women and 383 men were available for the matching procedure.

Propensity Score Matching (PSM; Rosenbaum & Rubin, 1984) was used to minimize differences between women and men on factors expected to be predictive of criminal behavior. PSM enables observational research as closely as possible to parallel the characteristics of randomized controlled studies by determining individual group membership conditioned upon a number of relevant observable variables (Austin, 2011). It has commonly been used to create gender-based groups that are equivalent on available variables that predict outcome (e.g., Alam et al., 2013). In this case we were seeking to equalize the levels of static-factor based risk across the two samples, before testing discrimination of the DRAOR. Given predictive accuracy may not be linearly related to recidivism in terms of discrimination or calibration (Andrews et al., 2012) this matching approach enabled us to compare the two genders in the same portion of the static risk range.

Eight matching variables were entered in a logistic regression model predicting gender (see Table 1). The logistic regression was significant, meaning that women and men differed statistically on the eight matching variables. Pseudo R-squared values ranged from 0.22 to 0.29 and the model correctly classified 71% of cases. Matched pairs (i.e., female and male) were generated using the nearest neighbor optimal matching method (Austin, 2011). A caliper was set for the women's propensity scores, meaning a man could be matched to a woman if his propensity score fell within +/- 0.008 of the woman's propensity score. The optimal matching method ensured the best-matched pairs were chosen. The matching procedure was conducted without replacement, meaning once a man had been matched to a woman he could not be matched again. The matching process was first carried out forwards: starting with women with the smallest propensity scores, and then in reverse: matching the women with the largest propensity scores first. This process further helped to ensure the finalized pairs were the optimal matches.

As a result of this procedure, each of 175 women was successfully matched to a male comparison (see Table 1). These women and men did not significantly differ from one another on any of the matching variables, confirming the success of the PSM procedure. Examination of those cases that were not matched showed that the women who were not matched had extremely low static risk scores ($M_{RoC*RoI}$ = .10, *SD* = .09), and that the men who were not matched were at the upper end of this sample, with medium static risk of imprisonment scores ($M_{\text{Roc*RoI}}$ = .44, *SD* = .18; see description below).[3]

[Insert Table 1 here]

Sample Characteristics

Together, the women and men in this sample had a mean age of just under 35 years (*SD* = 11 years) and an average community sentence length of 276 days (*SD* = 78 days). The largest proportion of people in the sample identified as New Zealand European (46%) or Māori[4] (43%), followed by a smaller proportion of Pacific Peoples (7%). The largest proportion of people (69%) committed a non-violent index offense (e.g., dishonesty and property offenses), approximately one quarter (24%) was convicted of a violent offense (e.g., assault and grievous bodily harm) and 7% was convicted of justice/administrative offenses (e.g., breach of intensive supervision orders). Overall, the resulting sample was low risk, based on the actuarial risk assessment scores routinely used in New Zealand to determine management of people on sentence in the correctional system (RoC*RoI *M* = 0.24, *SD* = 0.18; see description below).

Measures

The Dynamic Risk Assessment for Offender Re-entry (DRAOR). The DRAOR was developed by Serin (2007) and the New Zealand adaptation is currently in its third version (Serin et al., 2012). The DRAOR is a structured dynamic risk assessment tool designed to facilitate the assessment of recidivism risk in the community in addition to guiding risk management and case planning of people on sentences (Yesberg & Polaschek, 2015).

The DRAOR comprises 19 items, divided into three subscales: two risk subscales (Stable and Acute), and one Protective subscale. The dynamic risk items included in the risk subscales were adapted from research on stable and acute risk factors for sexual offending (Hanson & Harris, 2000) to relate to general and violent reoffending. The acute risk factors—substance abuse, anger/hostility, opportunity/access to victims, negative mood, employment, and interpersonal relationships—were developed to be proximal indicators of risk state (Douglas & Skeem, 2005), while the stable risk factors—peer associations, attitudes towards authority, impulse control, problem-solving, sense of entitlement, and attachment with others—represent criminogenic needs (Andrews & Bonta, 2010). The protective factors are internal strengths or external assets of the individual that are proposed to mitigate an person's risk of reoffending (Tamatea & Wilson, 2009); they are responsive to advice, prosocial identity, high expectations, costs/benefits, social supports, and social control. The protective factors were incorporated from the desistance literature because of the increasing evidence of their correlation with parole and treatment success.

Each DRAOR item is scored on a three-point scale (0, 1, and 2). For the Stable and Acute subscales a score of 0 indicates the item is not considered to be a problem for a given person, while a score of 2 is allocated when the item is presenting a considerable risk for reoffending (i.e., "definite problem"; Tamatea & Wilson, 2009). A score of 1 for an item is given if it is a possible problem, or the probation officer is uncertain of a score due to mixed or incomplete evidence (Tamatea & Wilson, 2009). The Stable and Acute risk subscales incorporate six and seven items respectively, and the highest score a person can be given on each is 12 and 14 respectively.

The Protective subscale is scored on the same 3-point scale, but a higher score signifies a greater degree of that factor. The Protective subscale includes six items and the highest score a person can be given is 12.

For each sample member, a DRAOR total score (TS) was manually calculated by summing the Acute and Stable risk subscale scores and subtracting the Protective subscale score. The TS can range from -12 to +26 and is an index of the person's risk corrected for his or her available protective factors. A higher DRAOR TS is indicative of higher risk of recidivism, because of the disproportionate presence of risk factors and relative absence of protective factors (Yesberg & Polaschek, 2015).

Probation officers score the DRAOR multiple times over a person's community sentence, at every non-trivial contact with the sentenced person. Continuous reassessment using the DRAOR provides up-to-date estimates of an individual's likelihood of reoffending. For this study, 'proximal' DRAOR scores were used: defined as the last assessment prior to a re-offense leading to conviction (for those who reoffended during their sentence). For women and men who were reconvicted for an offense committed after their sentence end date, or who were not reconvicted, the proximal DRAOR score was their last available assessment during the sentence. Paired samples t-tests showed there were significant differences between the first and last (or proximal) DRAOR TS and the Stable, Acute and Protective subscales, suggesting reassessment does provide a more up-to-date assessment of risk.[5] Average difference scores were 4.6 for the DRAOR total score, 2.0 for the Stable subscale, 2.1 for the Acute subscale, and -2.0 for the Protective subscale. Inter-rater reliability was not available for the DRAOR scores used in this study because they were completed by probation officers during routine practice and extracted retrospectively for this study.

The Risk of Re-conviction X Risk of Imprisonment. The RoC*RoI is an actuarial risk assessment scale based on static risk factors. It was developed in New Zealand and validated on two independent samples drawn from all those convicted of an imprisonable offense: each sample comprised 24,000 people (Bakker, Riley, & O'Malley, 1999). The RoC*RoI score is the probability that in the next five years the person will be reconvicted for an offense resulting in an imprisonment sentence. The static factors incorporated into the RoC*RoI algorithm consist of criminal history variables such as number of previous convictions and demographic factors (e.g., age; Bakker, et al., 1999). Analysis during development showed the RoC*RoI had moderate to high predictive validity (AUC = .76; Bakker, et al., 1999). More recent analysis has shown the RoC*RoI to have good predictive validity over three years post-release (Nadesu, 2007).

Recidivism. Recidivism data were extracted from the New Zealand National Conviction Records Database on 13 June 2014. Recidivism was defined as any new criminal reconviction, excluding convictions for breaches of community supervision, in the follow-up period (coded dichotomously; 0 = no, 1 = yes). The follow-up period or 'time-at-risk' consisted of the length of the person's community supervision sentence in addition to the length of time in the community post-sentence until the date of recidivism data extraction. Thus, time-at-risk was different for each person. The average time-at-risk for women was 804 days (SD = 199, range 382 to 1186) and for men, 802 days (SD = 218, range 374 to 1194). There was no significant difference between women and men's time-at-risk, $t(350) = 0.07$, $p = .941$.

For each person, survival time was calculated. For those who were reconvicted, survival time was the number of days between sentence start and the date on which the first offense occurred; for non-recidivists, survival time was days from sentence start until the date of data extraction.

Results

Some overall results will first be presented before addressing the three primary research questions. The mean subscale scores, in addition to the DRAOR total scores (TS), for women and men are presented in Table 2. Interestingly, despite being matched on static risk, the women scored significantly higher on the Acute and Stable subscales and significantly lower on the Protective subscale compared to their matched male counterparts. As a result, women had significantly higher DRAOR total scores compared to men.

[Insert Table 2 here]

Consistent with the differences in DRAOR scores identified above, with women being rated by their probation officers as significantly more 'risky' than men, women also had a higher recidivism rate. Thirty-six percent of women (n = 63) and 25% of men (n = 43) were reconvicted in the follow-up period. Kaplan-Meier survival analysis revealed this difference was statistically significant, Tarone-Ware statistic of equality = 4.77, df = 1, p = .029. As seen in Figure 1, women were reconvicted for new offenses faster and more often than their male counterparts.

[Insert Figure 1 here]

Research Question 1: Does the DRAOR Predict Reconviction with Equal Accuracy for Women and Men?

To first determine whether the DRAOR predicted reconvictions for women and men, univariate Cox regression survival analyses were conducted. Models were run separately for each sample. The DRAOR TS was entered as the independent variable, the criterion variable was the dichotomous reconviction variable, and the time variable was survival days. As shown in Table 3, the DRAOR TS significantly predicted reconvictions for women, $X^2(1, N = 175) = 10.59$, p = .003, and for men, $X^2(1, N = 175) = 10.48$, p = .001. For women, the hazard ratio was 1.08, meaning a one-point increase on the DRAOR TS was associated with an 8%

increase in the hazard of reconviction. For men, the same one-point increase was associated with a 10% increase in the hazard of reconviction.

Area Under the Receiver Operating Curve (AUC) values were also produced to facilitate the direct comparison of the DRAOR's predictive accuracy across gender. AUCs were used because they are robust to differences in the base rates of reconvictions. As seen in Table 3, the AUC values for women and men were significant, at .64 and .63 respectively. Using the widely-accepted interpretations of AUCs promulgated by Rice and Harris (2005), these magnitudes would be described as medium. The values indicate the DRAOR was able to differentiate between recidivists and non-recidivists equally for both genders. The confidence intervals associated with each gender's AUC values substantially overlapped, suggesting predictive accuracy was equivalent for women and men.

[Insert Table 3 here]

Research Question 2*:* Are There Differences in Which Subscales Underpin Predictive Validity for Women vs. Men?

Next, we explored whether the DRAOR subscale scores in combination predicted reconvictions, in addition to examining the differential predictive power of the individual subscales by gender. Identical to research question one, separate Cox regression models were run for each sample. For each model, the three subscales of the DRAOR were entered together as the independent variables; the criterion and time variables were identical to those of research question 1.

The multivariate model significantly predicted reconvictions for women, $X^2(3, N = 175) = 12.05$, $p = .007$, and men, $X^2(3, N = 175) = 12.39$, $p = .006$. Table 4 presents the unique contribution of each subscale to the respective models. For women, no subscale contributed uniquely to the model. For men, the Acute subscale contributed uniquely to the model; it predicted reconvictions even when we had already taken into account the contribution of the other subscales, suggesting it had the most influence on the ability of the DRAOR to predict recidivism in men. Both multivariate models were supported by significant AUC values of .64, indicating that after taking into account the unequal base rates of reconviction, the models had comparable moderate discriminative accuracy.

The failure of individual scales to demonstrate unique predictive validity in these multivariate analyses is likely a result of the high correlations between them, indicating substantial shared variance. For women these were: r = .468 between Stable and Acute, r = .358 between Acute and Protective, and r= -.615 between Stable and Protective subscales. The same correlations for men were respectively r = .450, -.424, and -.60. To further support this point, we also examined the ability of each of the subscales alone to predict recidivism for women and for men. For women, the Stable subscale alone significantly predicted recidivism: $X^2(1, N = 202) = 8.75$, $p = .003$, AUC = .63. For men, the corresponding result was $X^2(1, N = 175) = 6.54$, $p = .011$, AUC = .59. Turning to the Acute subscale alone, the Cox regression model for women was non-significant: $X^2(1, N = 175) = 3.10$, $p = .078$, AUC = .58. For men the model was statistically significant: $X^2(1, N = 175) = 9.95$, $p = .002$, AUC = .62, but the confidence intervals for men and women still overlapped for this pair of analyses. Lastly, for the Protective subscale, the models were significant for women, $X^2(1, N = 175) = 10.46$, p =.001, AUC = .64, and for men, $X^2(1, N = 175) = 5.17$, $p = .023$, AUC = .59.

[Insert Table 4 here]

Research Question 3: Does the DRAOR Provide Incremental Predictive Validity Above a Static-Factor Based Risk Estimate (the RoI*RoI) For Women and For Men?

The final research question asked whether the DRAOR contributed incrementally to the predictive validity of the static risk scale: the RoC*RoI. As with previous analyses separate Cox regression models were run for women and men. Models were run using the DRAOR TS, and the individual subscales. For each model, the RoC*RoI score was entered in the first block and the DRAOR was added in the second. As shown in Table 5, the RoC*RoI, on its own, significantly predicted reconvictions for women, $X^2(1, N = 175) = 11.34$, $p = .001$, and for men, $X^2(1, N = 175) = 20.29$, $p < .001$. The addition of the DRAOR TS in block 2 was significant for women, $X^2(2, N = 175) = 18.33$ $p < .001$, and men, $X^2(2, N = 175) = 27.84$, $p < .001$. In other words, for women and men, recidivism prediction was significantly improved by the addition of the DRAOR TS to the RoC*RoI.

Looking next at the incremental predictive validity of the individual subscales, the addition of the Acute subscale was significant for men, $X^2(2, N = 175) = 6.26$ $p = .012$, but not for women, $X^2(2, N = 175) = 1.69$ $p = .193$. This finding is consistent with the previous analyses showing the superior predictive validity of the Acute subscale for men. Conversely, the inclusion of the Stable subscale was significant for women, $X^2(2, N = 175) = 4.97$ $p = .026$, but not for men, $X^2(2, N = 175) = 3.23$ $p = .072$. Lastly, for both women and men, the Protective subscale added significantly to recidivism prediction over and above the RoC*RoI, women: $X^2(2, N = 175) = 9.02$ $p = .003$, men: $X^2(2, N = 175) = 3.92$ $p = .048$.

[Insert Table 5 here]

Discussion

How does a gender-neutral risk assessment tool compare in the prediction of recidivism for women and for men? This paper was the first to compare the predictive validity of the DRAOR with case-matched samples of community-sentenced women and men. Cox regression survival analyses with DRAOR scores that were as proximate to reconviction as was feasible significantly predicted reconviction for both genders. Furthermore, the scale had equivalent, moderate predictive accuracy for women and men as indicated by AUC values of .64 and .63 respectively.

These findings lend support to the view that scales developed for men are not inevitably unsuitable for use with women, and are consistent with previous research (e.g., Manchak, Skeem, Douglas, & Siranosian, 2009; Pusch & Holtfreter, 2018; Smith et al., 2009). The study builds on previous New Zealand research using the DRAOR (Yesberg et al., 2015) by using a more rigorous method to match the samples, and by using a lower risk, community-sentenced sample. By using equal numbers of women and men, and case matching sample members, it also improves on previous research that included disproportionately smaller numbers of women compared to men (e.g., Manchak et al., 2009).

Examination of the predictive validity of the three DRAOR subscales found notable differences by gender. For women, when examined in combination, no one subscale uniquely predicted reconvictions; but for men, the Acute subscale independently predicted reconvictions, suggesting it has a particularly important role in proximal prediction of recidivism for men, which is consistent with the theorized importance of acute factors (Hanson & Harris, 2000). These findings require replication, but suggest there may be underlying differences in the way DRAOR subscales function for men and women. The findings are consistent with previous research showing different subscales or domains within gender-neutral tools predict differently across gender (e.g., Olver et al., 2014). However, in contrast to our previous study which found the Acute subscale was more predictive for women than men (Yesberg et al., 2015), here we find it to be most predictive for men. Univariate Cox regressions, looking at each subscale individually, showed the Stable and Protective subscales significantly predicted recidivism for both genders, while the Acute subscale on its own was only predictive for men. Previous research has found protective factors may be more important for women than for men (Benda, 2005; McCoy & Miller, 2013): a finding that was not replicated here.[6]

Although the findings of this research provide some support for the general premise of gender neutrality, they do not necessarily suggest the DRAOR, or gender-neutral tools more broadly, are the best tools for use with women on community sentences. The field of correctional psychology is littered with risk assessment scales, and many have relatively limited empirical evidence of reliability and validity, particularly across jurisdictions, ethnicities and gender. Finding that the gender-neutrality of a tool is empirically supported is important from a practical as well as a theoretical perspective.

The research did not examine proposed gender-informed risk factors, and the observed differences in the predictive validity for women vs. men of the individual DRAOR subscales in combination, suggest the DRAOR is not operating in exactly the same way across gender. The predictive validity of the combined scales is moderate, and leaves room for improvement in prediction. It is entirely possible the inclusion of more gender-responsive risk factors may add incremental predictive validity above the DRAOR with women. Previous research has shown a series of gender-informed supplements may enhance the predictive power of an existing gender-neutral model (Van Voorhis et al., 2010). Nevertheless, because current practice with the DRAOR is predicated on the assumption of gender neutrality, the findings of this study are somewhat reassuring from this perspective.

Consistent with previous research showing the superior predictive ability of tools that incorporate dynamic risk factors (e.g., Olver, et al., 2007), the DRAOR made an incremental contribution to the prediction of recidivism over and above a static risk estimate, for both women and men. A dynamic risk score based on more up-to-date information added unique statistical variance in the prediction of recidivism. The DRAOR was designed to be useful in monitoring the day to day risk status of a person being managed on a community sentence, and to suggest points and areas in which to intervene to reduce risk. Finding that the proximal or most recent DRAOR scores predict recidivism over and above a static estimate (Table 5 Block 2) suggests its items are relevant to risk status.

Each case in this sample was closely matched on static-factor-estimated risk—using the RoC*RoI which is an established scale for use in assessing criminal risk for women and men in New Zealand—along with seven other variables predictive of recidivism, thus we had every reason to expect similar DRAOR scores and similar recidivism rates for each gender. Yet we found that, in comparison with men, women were reconvicted more often and more quickly (Figure 1), and further, their proximal DRAOR scores were significantly higher on all subscales (Table 2). Thus, although predictive validity was equivalent, one possible implication of these findings is that the RoC*RoI is not equivalently calibrated for women and men; at least in this low-risk range. Alternative explanations exist of course, including the possibility that women's risk-related needs are not being managed as actively as with men with the same level of RoC*RoI, leading to women's higher proximal DRAOR Acute and Stable scores, and lower Protective scores seen in Table 2. Another possibility, that women are being overscored on the DRAOR, would seem not to be supported here given the higher rates of reconviction for women.

Limitations and Future Directions

There are several key limitations to this study. Firstly, due to the samples available for matching—which consisted of a group of women who were, on average, at lower risk for serious recidivism than their male counterparts—the final sample evaluated were low-risk, as assessed by a static risk tool: men at higher risk were unable to be successfully matched to women and thus were removed from the research sample, and very women at very low risk similarly were removed because there were no men available with equivalent risk scores. We chose this approach for ease of interpretation of the results. For instance, had we instead used the propensity matching variable as a statistical control for the whole of both genders, it would not have been easy to see whether DRAOR scores were equivalent for those with equivalent RoC*RoIs.

Our selection process created a slightly artificial sample, although we would argue that preferentially attending to the higher risk end of the women's sample distribution has practical relevance. Those who were excluded are sufficiently low risk as to not justify active management of risk at all, in accordance with the Risk principle (Bonta & Andrews, 2017). On the other hand, our inability to match a number of very low risk women with similar men begs the question of whether women are being over-sentenced (e.g., being given a more intensive sentence when a man would simply be fined). A related possibility is that women have higher welfare needs than equivalent men, as other researchers have suggested, and the supervision sentence is imposed to provide an opportunity to better support the woman in this regard. Regardless, to fully match the women's sample at this low risk end, we would need to have oversampled for very low risk men. We were also not able to compare the performance of the DRAOR in the matched sample with the 161 women and 208 men who were not matched. Therefore, the generalizability of these findings to all community sentenced women (particularly those at the lower end of the risk scale) cannot be determined from this study and should be an area for future research.

Compared to past examinations with men, the DRAOR performed more poorly in this research. An explanation for this finding is the narrow range and low scores on static risk in this sample of men; previous evaluations of the DRAOR with men in New Zealand have included high-risk samples (Yesberg & Polaschek, 2015), or sampled a full range of risk levels (Lloyd, Hanson, Richards, & Serin, 2019). Because this is the first study that has examined the DRAOR with people serving community-based sentences in New Zealand, its interpretation requires caution and replication is needed.

The full value of the DRAOR, and dynamic risk scales more generally, resides in their ability to identify areas of need for people serving sentences for criminal behavior (Manchak et al., 2009), and to demonstrate that changes in these needs are associated with changes in recidivism outcomes. It could be argued that a further limitation of this study is that we did not use the DRAOR in the way it was intended. In other words, we used a dynamic risk instrument in a static way by only considering assessments made at one time point. Although we selected the measurement most proximal to reoffending that led to reconviction—which improves on previous research (e.g., Yesberg et al., 2015)—future research should measure change in DRAOR scores over time to establish whether DRAOR scales are relevant to risk change with community-sentenced people, especially women. Furthermore, this study was limited by relying on one measure of reoffending: any new conviction for a criminal offense. Future research should explore the ability of the DRAOR to predict alternative indices of recidivism (e.g., arrests, breaches, self-reported offending). A previous study with men on parole after undertaking prison-based intensive treatment vs. minimal treatment showed improvements in DRAOR scores over the course of parole, but the relationship of these changes to recidivism was not investigated (Polaschek & Yesberg, 2018). Two studies have now investigated whether updating DRAOR scores improves prediction over earlier scores, but both were with predominantly male parolee samples (Davies, 2018; Lloyd et al., 2019).

The main focus of this study was on empirically establishing practical utility. But the most energetic challenges to the use with women of scales developed for use with men come from a theoretical perspective. In this vein, a key direction for future research should be to continue to evaluate the performance of gender-informed factors in addition to a gender-neutral scale or model, with matched samples of women and men. Such an approach allows for the examination of gender-responsive factors with better methodological rigor (see for e.g., Bell, 2014). The male comparison group would enable direct appraisal of whether proposed gender-specific risk factors provide incremental predictive validity and, importantly, whether they do so uniquely for women as theorized. Promising gender-responsive factors include victimization, mental health difficulties and potentially a composite of the aforementioned factors in addition to substance abuse. Such a composite would tap into the unique cumulative impact that gender-responsive advocates (e.g., Hannah-Moffat, 2009) propose negative life experiences have on female criminality.

Conclusion

The present research was the first empirical study to examine the predictive validity of the DRAOR with matched samples of community-sentenced women and men. Based on the present research and the previous validation with men on parole (Yesberg et al., 2015) the DRAOR appears to be a valid risk assessment tool for the prediction of reconviction in women and men. However, it is important that future research continues to replicate research in this domain, as well as continuing to explore proposed gender-responsive factors to ensure women's levels of risk and types of need are being accurately captured.

References

Alam, M., Lee, V. V., Elayda, M. A., Shahzad, S. A., Yang, E. Y., Nambi, V., ... & Ramanathan, K. B. (2013). Association of gender with morbidity and mortality after isolated coronary artery bypass grafting: A propensity score matched analysis. *International Journal of Cardiology, 167*, 180-184. doi:10.1016/j.ijcard.2011.12.047

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

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

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

Baglivio, M. T., & Jackowski, K. (2013). Examining the validity of a juvenile offending risk assessment instrument across gender and race/ethnicity. *Youth Violence and Juvenile Justice, 11*, 26-43. doi:10.1177/1541204012440107

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

Bell, V. R. (2014). *Gender-responsive risk assessment in corrections*. El Paso, TX: LFB Scholarly Publishing.

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

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

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

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

Bloom, B., Owen, B. A., & Covington, S. (2003). *Gender-responsive strategies: Research, practice, and guiding principles for women offenders.* Washington, DC: US Department of Justice, National Institute of Corrections.

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

Brennan, T., Breitenbach, M., Dieterich, W., Salisbury, E. J., & Van Voorhis, P. (2012). Women's pathways to serious and habitual crime: A person-centered analysis incorporating gender responsive factors. *Criminal Justice and Behavior*, *39*, 1481-1508. doi:10.1177/0093854812456777

Carson, E. A. (2018). *Prisoners in 2016.* Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

Chadwick, N. (2014). *Validating the Dynamic Risk Assessment for Offender Re-entry (DRAOR) in a sample of U.S. probationers and parolees.* (Unpublished Doctoral thesis). Carleton University, Ottawa, Canada.

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

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

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

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

Department of Corrections. (2017a). *Briefing to the Incoming Minister*. Wellington, New Zealand: Department of Corrections.

Department of Corrections. (2017b). *Prison facts and statistics - September 2017.* Wellington, New Zealand: Department of Corrections. *Retrieved from:https://www.corrections.govt.nz/resources/research_and_statistics/quarterly_prison_statistics/prison_stats_september_2017.html*

Dickson, S. R., Polaschek, D. L., & Casey, A. R. (2013). Can the quality of high-risk violent prisoners' release plans predict recidivism following intensive rehabilitation? A comparison with risk assessment instruments. *Psychology, Crime & Law, 19*, 371-389. doi:10.1080/1068316X.2011.640634

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

Fortune, C.-A., & Ward, T. (2017). Problems in protective factor research and practice. *Aggression and Violent Behavior, 32*, 1-3. doi:10.1016/j.avb.2016.12.008

Geraghty, K. A., & Woodhams, J. (2015). The predictive validity of risk assessment tools for female offenders: A systematic review. *Aggression and Violent Behavior, 21,* 25-38. doi:10.1016/j.avb.2015.01.002

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30. doi:10.1037/1040-3590.12.1.19

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

Hannah-Moffat, K. (2009). Gridlock or mutability: Reconsidering "gender" and risk assessment. *Criminology & Public Policy*, *8*, 209-219.

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

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

Kubiak, S. P., Kim, W. J., Fedock, G., & Bybee, D. (2012). Assessing short-term outcomes of an intervention for women convicted of violent crimes. *Journal of the Society for Social Work and Research, 3,* 197-212. doi:10.5243/jsswr.2012.13

Lloyd, C. D., Hanson, R. K., Richards, D. K., & Serin, R. C. (2019). *Reassessment improves prediction of criminal recidivism: A prospective study of 3,421 individuals in New Zealand.* Manuscript under review.

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

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*, 425-442. doi:10.1177/0093854809333058

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

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

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

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*, 156-176. doi:10.1037/a0035080

Olver, M. E., Wong, S. C., 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*, 318-329. doi:10.1037/1040-3590.19.3.318

Olver, M. E., & Jung, S. (2017). Incremental prediction of intimate partner violence. *Law and Human Behavior*, *41*, 440-453. doi:10.1037/lhb0000251

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

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

Pusch, N., & Holtfreter, K. (2018). Gender and risk assessment in juvenile offenders: A meta-analysis. *Criminal Justice and Behavior, 45,* 56-81. doi:10.1177/0093854817721720

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

Rettinger, L. J., & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior, 37,* 29-46. doi:10.1177/0093854809349438

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

Roe-Sepowitz, D. E., Bedard, L. E., Pate, K. N., & Hedberg, E. C. (2014). Esuba: A psychoeducational group for incarcerated survivors of abuse. *International Journal of Offender and Comparative Criminology, 58*, 190-208. doi:10.1177/ 0306624X12465410

Rosenbaum, P. R., & Rubin, D., B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79,* 516-524.

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

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

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

Smith, P., Cullen, F., & Latessa, E. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy, 8,* 183–208.

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

Statistics New Zealand (2012) *New Zealand's Prison Population.* Available from: http://www.stats.govt.nz (accessed 16 June 2014).

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

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

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

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

Van Voorhis, P. (2012). On behalf of women offenders: Women's place in the science of evidence-based practice. *Criminology and Public Policy, 11*, 111-145.

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,* 261-288. doi:10.1177/0093854809357442

Vitopoulos, N. A., Peterson-Badali, M., & Skilling, T. A. (2012). The relationship between matching service to criminogenic need and recidivism in male and female youth: Examining the RNR principles in practice. *Criminal Justice & Behavior, 39*, 1025-1041. doi:10.1177/0093854812442895

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

Wattanaporn, K.A., & Holtfreter, K. (2014). The impact of feminist pathways research to gender-responsive policy and practice. *Feminist Criminology, 9*, 191-207. doi:10.1177/1557085113519491

Wright, E. M., Salisbury, E. J., & Van Voorhis, P. (2007). Predicting the prison misconducts of women offenders: The importance of gender-responsive needs. *Journal of Contemporary Criminal Justice, 23*, 310-340. doi:10.1177/1043986207309595

Yesberg, J. A., & Polaschek, D. L. L. (2015). Assessing dynamic risk and protective factors in the community: Examining the validity of the Dynamic Risk Assessment for Offender Re-entry. *Psychology, Crime, & Law, 21,* 88-99. doi:10.1080/1068316X.2014.935775

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

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

Endnotes

1. Daly described five criminal trajectories that women follow: (1) Street women (2) Harmed and harming women, (3) Battered women, (4) Drug-connected women and (5) Economically motivated women. Four of the five pathways are considered gendered because they are hallmarked by gender-responsive factors. However, the economically motivated pathway is non-gendered because it is also seen in male offenders.
2. This manuscript reports results from a more extended study in which multiple DRAOR scores were used; hence the exclusion criteria. See the Measures section for information on when the DRAOR is scored in routine practice. Any of these exclusions is likely to indicate poor implementation integrity on the part of the assessors.
3. The RoC*RoI scores are categorized in this study as per the New Zealand Parole Board standard categories: 'Low' ($0 \geq 0.25$), 'Low-moderate' ($0.25 \geq 0.5$), 'Moderate' ($0.5 \geq 0.7$), 'High' ($0.7 \geq 0.8$), and 'Very high' (0.8+).
4. Indigenous people of Aotearoa New Zealand
5. DRAOR TS: $t(349) = 14.63$, $p < .001$; Stable: $t(349) = 11.12$, $p < .001$; Acute: $t(349) = 11.70$, $p < .001$; Protective: $t(349) = -10.93$, $p < .001$
6. Although in the multivariate Cox regressions the Protective subscale was nearing significance for women: the upper confidence interval in Table 4 was exactly 1.00.

Table 1

Matching variables and equivalence tests for the final samples

	Women (n = 175)	Men (n = 175)	Test of Equivalence	
	n (%)	n (%)	Chi-Square Analysis of Variance	Phi (Φ)
Ethnicity				
Māori	73 (41.7)	76 (43.4)		
European	84 (48.0)	78 (44.6)	$\chi^2(3, n = 350) = 0.51, p = .916$	0.04
Pacific Peoples	12 (6.9)	14 (8.0)		
Other	6 (3.4)	7 (4.0)		
Index Offense				
Non-violent	119 (68.0)	121 (69.1)		
Violent	42 (24.0)	42 (24.0)	$\chi^2(2, n = 350) = 0.17, p = .918$	0.02
Justice/admin	14 (8.0)	12 (6.9)		
	M (SD)	*M (SD)*	Independent Samples t test	Cohen's d
Age (years)	34.86 (11.16)	34.06 (10.92)	$t(350) = 0.67, p = .501$	0.07
Sentence length (days)	276.03 (74.37)	275.05 (81.21)	$t(350) = 0.12, p = .906$	0.01
RoC*RoI score[a]	.25 (.19)	.24 (.17)	$t(350) = 0.19, p = .851$	0.02
# previous convictions	16.90 (15.53)	16.89 (18.87)	$t(350) = 0.00, p = .998$	0.00
# previous violent convictions	1.12 (1.49)	1.19 (1.57)	$t(350) = -0.42, p = .675$	-0.06
# previous imprisonments	.58 (1.57)	.70 (1.93)	$t(350) = -0.61, p = .543$	-0.07

[a] Values reported do *not* assume equal variance.

Table 2

Proximal DRAOR subscale and total scores for women and men

	Women (n = 175)	Men (n = 175)	Analysis of Variance	Cohen's d
	M (SD)	M (SD)		
Acute	4.07 (2.24)	3.05 (1.94)	$F(3, 350) = 4.52, p < .001$	0.49
Stable	4.69 (2.51)	3.96 (2.42)	$F(3, 350) = 0.54, p = .006$	0.45
Protective	7.27 (2.38)	8.03 (2.27)	$F(3, 350) = 0.27, p = .002$	-0.33
DRAOR TS[a]	1.49 (5.78)	-1.03 (5.41)	$F(3, 350) = 0.55, p < .001$	0.45

[a] DRAOR TS is the total score (Stable + Acute – Protective).

Table 3

Cox regression survival analysis: Proximal DRAOR total score predicting reconviction

	Women (n = 175)				Men (n = 175)			
	β (SE)	Wald (df = 1)	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald (df = 1)	Hazard Ratio [95% CI]	AUC [95% CI]
DRAOR TS	0.07 (0.02)	10.61**	1.08 [1.03, 1.13]	.64** [.56, .76]	0.09 (0.03)	10.48**	1.10 [1.04, 1.16]	.63** [.53, .73]

* p<.05, ** p<.01

Table 4

Multivariate model of the proximal DRAOR subscales predicting reconvictions

	Women (n = 175)				Men (n = 175)			
Subscale	β (SE)	Wald (df = 1)	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald (df = 1)	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.01 (0.07)	0.02	1.01 [0.89, 1.15]		0.18 (0.08)	4.53*	1.19 [1.01, 1.40]	
Stable	0.08 (0.07)	1.46	1.09 [0.95, 1.24]	.64** [.55, .72]	0.09 (0.08)	1.22	1.09 [0.94, 1.27]	.64** [.54, .74]
Protective	-0.13 (0.07)	3.65	0.88 [0.77, 1.00]		-0.04 (0.09)	0.19	0.96 [0.82, 1.14]	

* p<.05, ** p<.01

Table 5

Incremental predictive validity of the proximal DRAOR total score

	Women (n = 175)			Men (n = 175)		
	β (*SE*)	Wald (*df* = 1)	Hazard Ratios [95% CI]	β (*SE*)	Wald (*df* = 1)	Hazard Ratios [95% CI]
Block 1						
RoC*RoI	2.04 (0.61)	10.98**	7.66 [2.30, 25.54]	3.29 (0.75)	19.10***	26.89 [6.15, 117.67]
Block 2						
RoC*RoI	1.72 (0.64)	7.36**	5.60 [1.61, 19.42]	2.86 (0.76)	14.27***	17.52 [3.97, 77.42]
DRAOR TS	0.06 (0.02)	7.40**	1.07 [1.01, 1.12]	0.07 (0.03)	6.54*	1.08 [1.02, 1.14]
Block 2						
RoC*RoI	1.93 (0.63)	9.49**	6.86 [2.02, 23.36]	2.91 (0.75)	15.19***	18.36 [4.25, 79.31]
Acute	0.07 (0.06)	1.69	1.08 [.96, 1.20]	0.18 (0.07)	6.26*	1.20 [1.04, 1.39]
Block 2						
RoC*RoI	1.68 (0.64)	6.88**	5.38 [1.53, 18.90]	2.99 (0.77)	15.14***	19.91 [4.41, 89.85]
Stable	0.11 (0.05)	4.97**	1.12 [1.01, 1.24]	0.12 (0.07)	3.23	1.12 [0.99, 1.28]
Block 2						
RoC*RoI	1.91 (0.63)	9.29**	6.76 [1.98, 23.12]	3.14 (0.75)	17.63***	22.98 [5.32, 99.28]
Protective	-0.17 (0.06)	9.02**	0.85 [0.76, 0.94]	-.13 (0.07)	3.92*	0.88 [0.77, 1.]

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

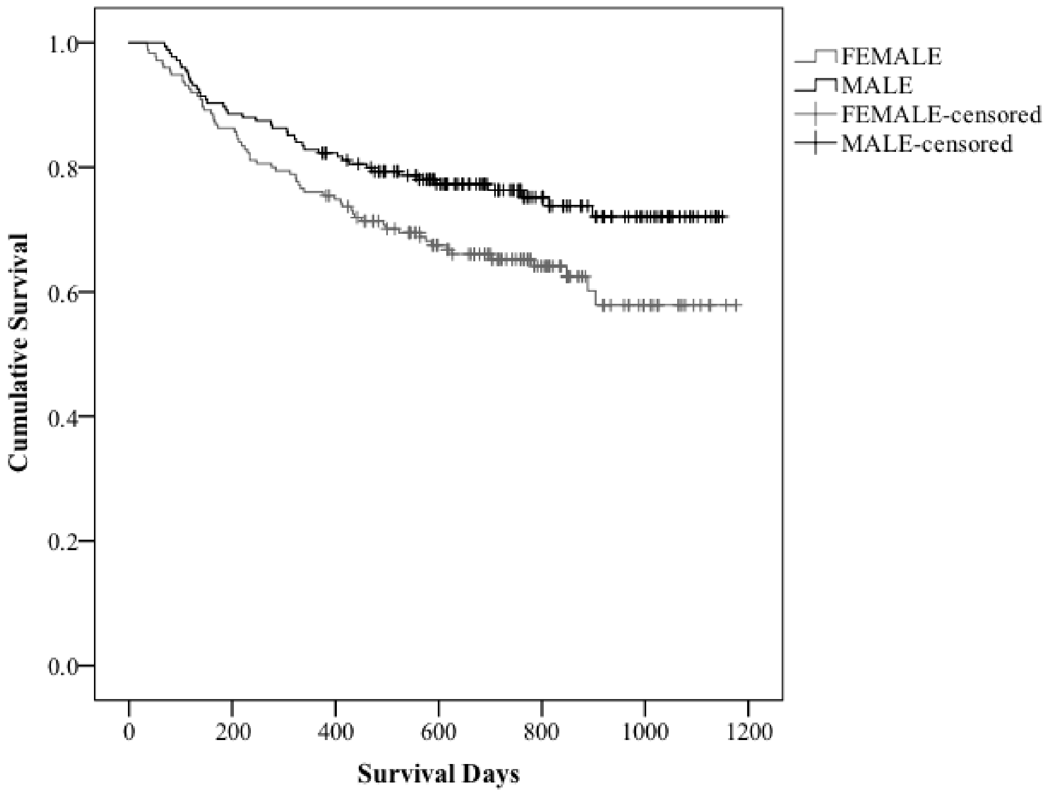

Figure 1. Survival curves of reconviction for women and men

Bio sketches

Jessica Scanlan, MSc, PGDipClinPsyc, is a clinical psychologist, currently working in the Regional Intellectual Disability Secure Service, with Capital and Coast District Health Board in Wellington, New Zealand. As a member of the multidisciplinary professional team she is tasked with the development and delivery of care and rehabilitation for persons with a diagnosed intellectual disability whom have committed imprisonable offences.

Julia Yesberg, PhD, is a Research Fellow in the Jill Dando Institute for Security and Crime Science at University College London. She is currently completing an ESRC-funded post-doctoral fellowship, exploring relationships between neighbourhood policing, collective efficacy, and violent crime. Her research interests include policing and public opinion, police use of force, serious violent crime, offender rehabilitation and reintegration, and risk assessment.

Clare-Ann Fortune, PhD, PGDipClinPsyc, is a Senior Lecturer in Clinical Forensic Psychology at the School of Psychology at Victoria University of Wellington, New Zealand. She is also registered with the New Zealand Psychologists Board in the Clinical scope of practice. Her research focuses on risk assessment, offender rehabilitation, and factors impacting the ability of young people to participate in the justice system. She has published papers on strength-based approaches to offender rehabilitation, the conceptualisation of risk and risk assessment.

Devon Polaschek, PhD, DipClinPsyc, MNZM, is a clinical psychologist, and professor in the School of Psychology, and the Director of the New Zealand Institute of Security and Crime Science, University of Waikato. Her research interests include understanding, predicting and preventing serious and violent reoffending, family violence, psychopathy, imprisonment, desistance, reintegration and parole.