Assessing Risk for Intimate Partner Violence: A Cross-Validation of the ODARA and DVRAG within a Sample of Incarcerated Offenders

by

Andrew L. Gray
BAH, Carleton University, 2010
BST, St. Lawrence College, 2007

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

Carleton University
Ottawa, Canada

©2012 Andrew L. Gray
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.
Abstract

This study was a cross-validation of the Ontario Domestic Assault Risk Assessment (ODARA) and the Domestic Violence Risk Appraisal Guide (DVRAG) in a sample of 94 offenders under federal jurisdiction in the Ontario region. Also included were the Psychopathy Checklist-Revised (PCL-R), the Statistical Information on Recidivism-Revised 1 (SIR-R1), and the Spousal Assault Risk Assessment (SARA). In a retrospective-prospective study design, offenders were followed for an average of 65.04 months yielding a base rate of 12.8% for intimate partner violence (IPV) recidivism. Statistical analyses revealed that the DVRAG and ODARA displayed high inter-rater reliability and that the two measures along with Factor 1 of the PCL-R generated the largest AUC values for IPV recidivism (AUC = .713, .712, and .685, respectively) relative to the PCL-R, SARA, and SIR-R1. Meta-analyses incorporating the current results are presented, as are discussions concerning the implications of utilizing these risk assessment measures with federal IPV offenders.
Dedication

For mom.
Acknowledgements

Once again, I would like to thank my supervisor, mentor, and friend Dr. Jeremy Mills for his ongoing guidance and support. Throughout the past four years, Jeremy has been instrumental in my continued personal, professional, and academic growth. Working with him has truly been an honour, and I will forever be indebted to him.

Many thanks to my co-supervisor Dr. Adelle Forth who provided me with ongoing support and encouragement (while also keeping me on track) throughout the completion of this thesis. I would also like to thank my committee members, Dr. Ralph Serin, Dr. Kevin Nunes, and Dr. Lara Karaian for their helpful comments. Additionally, completion of this thesis would not have been possible without the help of Carolyn Bourgeois whose assistance in data collection is gratefully appreciated. Finally, I would also like to acknowledge Dr. N. Zoe Hilton for her patience and assistance in answering my questions regarding the ODARA, and to the Ontario Graduate Scholarship (OGS) program for funding this project.

To my family, I am truly grateful. Your ongoing love and support has made all of this possible. These past two years have been among the most rewarding and most difficult years of my life thus far. Dad, against all odds, you have continued to support me throughout this journey and have made me laugh when I needed it most. Mom, your love and support throughout my life has provided me with the strength and courage to achieve my dreams and continually better myself. The courage and strength you have shown over the past year and a half is remarkable. And, as always, you never cease to amaze me. To my big brother, Greg, and his beautiful family (Krista, Kenzie, Rowan, Mason, and Ethan) and my two wonderful grandmothers, Dorothy and Christina, all of
your love and support helped to make this thesis a reality. I would also like to thank Cheryl for the many encouraging conversations, and to Michael for his ongoing interest and input. And to Elin, I will be forever grateful for your unwavering love and support throughout the completion of this project...you truly believed in me every step of the way. Despite the tragedy and heartache we have all endured, we continue to grow stronger as a family. I love you all, and I hope that you are all as proud of me as I am of you.

Finally, to all my friends who have supported me along this journey, I thank you. I also want to give a shout out to Ed. I hope that this brings you one step closer to winning your bet.
Table of Contents

Abstract ......................................................................................................................................... ii

Dedication ................................................................................................................................... iii

Acknowledgments ....................................................................................................................... iv

List of Tables ............................................................................................................................ viii

List of Appendices ....................................................................................................................... x

Introduction ...................................................................................................................................1

Violence Risk Assessment ..........................................................................................................3

Approaches to the Assessment of Violence Risk ...................................................................... 5

Psychopathy and Violence Risk Assessment ............................................................................. 14

Risk Assessment for IPV ........................................................................................................... 19

Development of the ODARA and DVRAG ............................................................................. 23

A Meta-Analytic Investigation into the Accuracy of Predicting IPV ....................................... 29

Moderator Analysis .................................................................................................................... 34

Study Purpose and Hypotheses ................................................................................................. 36

Hypotheses ................................................................................................................................ 37

Power Analysis and Proposed Sample Size ............................................................................. 38

Method ........................................................................................................................................ 42

Participants ................................................................................................................................. 42

Measures and Recidivism Variables .......................................................................................... 42

The Domestic Violence Risk Appraisal Guide ........................................................................ 42

The Ontario Domestic Assault Risk Assessment ...................................................................... 43
List of Tables

Table 1: Structured Components for Three Common Violence Risk Assessment Processes.......................................................................................................................................6

Table 2: Items from the Ontario Domestic Assault Risk Assessment (ODARA) and Domestic Violence Risk Appraisal Guide (DVRAG) ..................................................................................................................26

Table 3: Weighted AUC Values for Five Risk Assessment Measures in Predicting IPV.32

Table 4: Q-Between Analyses for Four Risk Assessment Measures in Predicting IPV ....34

Table 5: Sample Characteristics ..............................................................................................53

Table 6: Item Endorsement and Corrected Item-Total Correlations ($r_{ni}$) for the ODARA and DVRAG ..................................................................................................................................................57

Table 7: Inter-Rater Reliability Analyses ..............................................................................59

Table 8: Correlations between the ODARA, DVRAG and the PCL-R, SARA, SIR-R1, and Victim Injury ..................................................................................................................................................61

Table 9: Accuracy of Five Risk Assessment Measures in Predicting Violent Recidivism and Any Recidivism ($N = 94$) ..................................................................................................................................................63

Table 10: Accuracy of Five Risk Assessment Measures in Predicting IPV Recidivism Outcomes ($N = 94$) ..................................................................................................................................................64

Table 11: Partial Correlations Controlling for Treatment Dosage .......................................67

Table 12: Hierarchical Cox Regression Survival Analysis for Violent and IPV Recidivism ..................................................................................................................................................71
Table 13: Updated Weighted AUC Values for Five Risk Assessment Measures in Predicting IPV ............................................................................................................................83

Table 14: Updated Q-Between Analyses for Four Risk Assessment Measures in Predicting IPV ............................................................................................................................84
ASSESSING RISK FOR INTIMATE PARTNER VIOLENCE

List of Appendices

Appendix A: AUC Meta-Analysis Syntax for SPSS ...........................................................106
Appendix B: AUC Power Analysis Syntax for SPSS ..........................................................108
Appendix C: Certificate of Ethics Approval .................................................................111
Appendix D: Research Application and Undertaking .........................................................112
Appendix E: ODARA Certification .......................................................................................114
Appendix F: Comparisons Between the Moderate and High Intensity FVPP ..........115
Appendix G: Correlational Analysis Between the PCL-R, SARA, and SIR-R1 ..........116
Appendix H: Accuracy of Five Risk Assessment Measures in Predicting Revocation of Conditional Release (N = 88) .................................................................117
Assessing Risk for Intimate Partner Violence: A Cross-Validation of the ODARA and DVRAG within a Sample of Incarcerated Offenders

Intimate partner violence (IPV)\(^1\) is a widespread global phenomenon which can manifest itself in the form of physical, sexual, psychological, verbal, and economic abuse (Canadian Resource Centre for Victims of Crime, n.d.). In response to the growing issue of IPV, the early 1980s brought about legislative change within the Canadian criminal justice system by way of the introduction and subsequent adoption of policy directives which now required perpetrators of IPV to be charged and prosecuted by both police and Crown prosecutors so long as "there were reasonable and probable grounds to believe that an offence had been committed" (i.e., "pro-charging" and "pro-prosecution" policies; Department of Justice Canada, 2003, p. 1). Subsequent to these policy changes, other specialized measures have since been introduced into the Canadian criminal justice system such as domestic violence courts, services and treatment programs, and civil legislation for victims. The implementation of these policies was related to the perceived inadequate responses to IPV previously exhibited by the criminal justice system (Brown, 2000). No longer would IPV be treated as a "private matter," rather it had become a "criminal matter" (Department of Justice Canada, 2003). With the adoption of these policies and the resulting systematic research on their effectiveness, there has been increased interest in relation to IPV (Kropp, 2004).

In the 2009 General Social Survey, approximately six percent of Canadians reported being victims of spousal violence, both physical and sexual, within the previous

\(^1\) In keeping with the current state of the literature (see Helmus & Bourgon, 2011) IPV was adopted as the appropriate nomenclature as it allows for a definition that incorporates individuals (male or female) who have experienced violence committed by someone whom they have or have had "an intimate, sexual relationship" regardless of their marriage or cohabitation status (Kropp, Hart, Webster, & Eaves, 1999, p. 1).
five years (Brennan, 2011). Of the 554 police-reported homicides committed in Canada during 2010, 89 were committed by an intimate partner (Mahony, 2011). Within the United States, estimates indicate that approximately 1.5 million women and 834,732 men experience sexual and/or physical violence perpetrated by an intimate partner annually (Tjaden & Theonnes, 2000). Interviewing 24,000 women ranging from 15 to 49-years of age across 10 countries, the World Health Organization (WHO) found that approximately 15 to 71 percent of women reported experiencing some form of physical or sexual IPV during their lifetimes (Garcia-Moreno, Heise, Jansen, Ellsberg, & Watts, 2005). The extreme variations in the percentage of women reporting IPV violence was considered to be potentially related to the socioeconomic status of the urban centre surveyed (e.g., Japan city reportedly had the lowest reported rates of prior and current IPV). According to Garcia-Moreno and colleagues this “suggests that women in industrialized nations may find it easier to leave abusive relationships” (p. 1282) which may account for the lower prevalence rate within Canada relative to other countries. Nonetheless, survey results such as these are a clear indication that IPV remains a highly prevalent and ongoing societal concern.

The purpose of this study was to investigate the reliability and validity of two actuarial risk assessment measures designed for IPV offenders: the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004; Hilton, Harris, & Rice, 2010) and the Domestic Violence Risk Appraisal Guide (DVRAG; Hilton et al., 2010; Hilton, Harris, Rice, Houghton, & Eke, 2008) an indepth actuarial measure which combines both the ODARA and the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 2003). This study allowed for comparative analyses between the ODARA, DVRAG, PCL-R, the
Violence Risk Assessment

Violence risk assessment can provide a framework for the appropriate management and treatment of violent offenders (Andrews & Bonta, 2010; Hart, 2008). However, effective management and treatment of violent offenders is predicated upon correctly estimating or assessing their level of risk. Mills, Kroner, and Morgan (2011, p. 1) define risk assessment as:

1) determining an individual’s level of risk (risk estimation), (2) identifying the salient risk factors that contribute to that risk, (3) identifying risk management strategies and considerations to manage or minimize that risk, and (4) communicating the risk information to the decision maker.

Within the context of clinical practice, many psychologists and other mental health professionals are frequently relied upon by forensic decision makers to conduct informed risk assessments (Mills et al., 2011). In Canada, the assessment of violence risk has become an integral component of the Canadian criminal justice system, driven by legislative changes within Canadian law that were adopted from civil commitment laws within the United States (Lyon, Hart, & Webster, 2001). Within their 2003 US survey, Tolman and Mullendore (2003) found that, while only 9% of surveyed clinicians identified as forensic psychologists, over half (53%) had conducted some form of court-related risk assessment. In contrast, 89% of surveyed diplomates, or board certified forensic psychologists, had conducted some form of court-related risk assessment. The
most common context for the risk assessments was within criminal sentencing (87% vs. 45% for diplomates and clinicians, respectively), whereas the least was IPV assessments (40% vs. 18% for diplomates and clinicians, respectively). In a more recent survey of 215 forensic psychologists with experience in forensic or court-ordered assessments, Viljoen, McLachlan, and Vincent (2010) found that the majority of their sample (n = 119) had conducted at least one violence risk assessment exclusively for either juvenile (n = 77) or adult offenders (n = 122) with an overall average number of 204.38 and 429.48 violence risk assessments being conducted throughout the course of their careers, respectively.

As of 2001, “dangerous to self and others” (in addition to “requirement of mental disorder”) has been adopted as one of the criterion for the legal standard for civil commitment across all Canadian jurisdictions (Douglas & Koch, 2001). The clinical assessment of violence risk within legal proceedings addressing indefinite incarceration has become a requirement in both Canada and the United States as evidenced by dangerous offender (DO) hearings and sexually violent predator (SVP) hearings, respectively (Boccaccini, Murrie, Caperton, & Hawes, 2009; Heilbrun, Ogloff, & Picarello, 1999). Indeed, according to Lyon and colleagues (2001, p. 322), “assessments of risk are mandated in some form or other at virtually every major junction of the criminal justice system.” Similarly, Lyon and colleagues note that within civil settings the assessment of violence risk has been mandated for the following areas: civil commitment, duty to warns/limitations to confidentiality, child protection, immigration, and employment/schools. As a result, mental health professionals conducting violence risk assessments must strike a delicate balance between risk management and public safety (Douglas & Skeem, 2005). With the ethical considerations that must be taken into
account (i.e., offender civil liberties vs. public safety), it is of no surprise that violence risk assessment has become a controversial practice and that the accuracy of these assessments must withstand strict empirical investigation (Mills et al., 2011).

**Approaches to the Assessment of Violence Risk**

Skeem and Monahan (2011) recently proposed that the approaches to the assessment of violence risk fall along a “continuum of rule-based structure” which incorporate none to all of the following components: identifying empirically valid risk factors, determining a method for measuring (or “scoring”), establishing a procedure for combining scores on the risk factors, and producing an estimate of violence risk. Table 1 is based on Skeem and Monahan’s conceptualization of the risk assessment process and provides a summary of the three most common processes for violence risk assessment. The two polar opposites occupying either end of the continuum are: clinical judgment and statistical/mechanical prediction, respectively (the latter of which has become more commonly referred to as actuarial; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954; Skeem & Monahan, 2011). Occupying the completely unstructured end of the continuum, clinical judgment reflects an unstructured prognostic process in which violence risk is generally determined by “psychological structures and dynamics” (Hanson, 2009, p. 173) with no a priori method for selecting, measuring, or combining risk factors (Hart, 2008; Hilton, Harris, & Rice, 2006; Skeem & Monahan, 2011). This form of risk assessment is now more commonly known as unstructured professional judgment and is considered a part of the professional judgment procedures (Hart, 1998, 2008).
Table 1
Structured Components for Three Common Violence Risk Assessment Processes

<table>
<thead>
<tr>
<th>Violence risk assessment process</th>
<th>Structured component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured professional judgment (UPJ)</td>
<td>None</td>
</tr>
<tr>
<td>Structured professional judgment (SPJ)</td>
<td>Identify risk factors</td>
</tr>
<tr>
<td></td>
<td>Measure risk factors</td>
</tr>
<tr>
<td>Actuarial</td>
<td>Identify risk factors</td>
</tr>
<tr>
<td></td>
<td>Measure risk factors</td>
</tr>
<tr>
<td></td>
<td>Combine risk factors</td>
</tr>
<tr>
<td></td>
<td>Produce final risk estimate</td>
</tr>
</tbody>
</table>

Detractors of unstructured professional judgment have described it using terms such as subjective, unscientific, unreliable, intuitive, and impressionistic (Meehl, 1954; Grove & Meehl, 1996). Within his influential work, Meehl (1954), in reviewing the existing empirical literature pertaining to the prediction of three forms of human behaviour (i.e., scholarly success, recidivism, and relapse of psychotic symptoms), concluded:

In spite of the defects and ambiguities present, let me emphasize the brute fact that we have here, depending upon one's standards for admission as relevant, from 16 to 20 studies involving a comparison of clinical and actuarial methods, in all but one of which the predictions made actuarially were either approximately equal or superior to those made by a clinician. (p. 119)

Comparative results such as these in which poor predictive accuracy has been found for unstructured professional judgment in comparison to actuarial methods are replete throughout the empirical literature (e.g., Gardner, Lidz, Mulvey, & Shaw, 1996;
Quinsey, Harris, Rice, & Cormier, 2006; Steadman, 1983) and have been reflected within a number of meta-analytic studies (e.g., Ægisdóttir et al., 2006; Grove et al., 2000; Hanson & Morton-Bourgon, 2009). Grove and Meehl (1996), reviewing results of the meta-analysis later published by Grove and colleagues (2000) declared that the use of unstructured professional judgment as a means of predicting human behaviour "is not only unscientific and irrational, it is unethical" (p. 320).

In contrast, purely actuarial risk assessment measures consist of all four of the components outlined by Skeem and Monahan (2011). While some inconsistencies exist within the literature as to the definition of actuarial risk assessment, Mills and colleagues (2011) define the procedure as "any [risk assessment measure] that has a structured scoring method and associates a statistical or probabilistic statement with the resulting score" (p. 3). The structured scoring method mentioned by Mills and colleagues incorporates three of the four components from Skeem and Monahan (i.e., identifying, measuring, and combining risk factors); however, it is the statistical or probabilistic statement that sets it apart from other forms of risk assessment (i.e., the production of a final risk estimate). By virtue of their nature, actuarial risk assessment measures are solely rooted within an empirical model (Hanson, 2009; Mills, 2005). Historically, actuarial risk assessment measures have predominantly been comprised of empirically derived static risk factors, or historical variables, that were known to be statistically related to recidivism (Hanson, 2009). Static risk factors remain relatively stable over time and are not amenable to treatment (e.g., criminal history variables such as age at first offence; Andrews & Bonta, 2010).
Within Canada, one of the first actuarial risk assessment measures to be developed was the General Statistical Information on Recidivism scale (GSIR; Nuffield, 1982). Initially, the GSIR was developed with a sample of 2,500 Canadian federal offenders released into the community between 1970 and 1972. Excluded from the development sample were offenders who had been transferred from provincial jurisdiction or who had revocation status. Of the original 2,500 offenders, follow-up data was available for a total of 2,475 offenders. To facilitate the development and subsequent validation of the risk assessment measure, Nuffield randomly split the sample in half. Recidivism was coded as any re-arrest for an indictable offence within a three year follow-up period. A total of 15 variables were found to be predictive of recidivism: current offence, age at admission, previous imprisonments, previous breach of parole or mandatory supervision, previous history of escape, security classification, age at first adult conviction, previous convictions for assault, marital status, interval at risk since last offence, number of dependents, aggregate sentence, previous convictions for violent sex offences, previous convictions for break and enter, and employment status at time of arrest for current offence. Statistical weights were developed for each item using increments of approximately five percent, a plus or minus 1 was applied depending on the direction that the variable deviated from the recidivism base-rate (known as the simple summation techniques; Nafekh & Motiuk, 2002). Variables which deviated from the base-rate by less than five percent received a weighting of zero. The 15 weighted variables were then combined to form a total risk score with positive scores representing an increased probability of re-offending. Risk categories ranging from “very good” to “poor” were developed to provide probability estimates regarding release failure within
the initial three years of release from federal custody. Revisions were later made to the GSIR to account for legislative changes and to reflect research findings pertaining to the presence of prior sexual offenses (Bonta, Harman, Hann, & Cormier, 1996; Nafekh & Motiuk, 2002). Moreover, scores on the measure were reversed such that positive scores were now reflective of increased probability of succeeding in the community. In light of the revisions made to the GSIR, it was renamed the Statistical Information on Recidivism-Revised 1 (SIR-R1; Nafekh & Motuik, 2002) scale. Although originally designed to predict general recidivism (i.e., any recidivism), meta-analytic studies have found the SIR-R1 to be a moderate predictor of violent recidivism among sex offender ($k = 3, N = 566$, mean weighted Cohen's $d = 0.80$, 95% CI = [0.61, 0.99]; Hanson & Morton-Bourgon, 2009) and non-sex offender samples ($k = 17, N = 5,618$, $Z^* = .22$, 95% CI = [.19, .25]; Campbell, French, & Gendreau, 2009).

Notwithstanding the accuracy of actuarial risk assessment measures in identifying offenders at high risk for recidivism, much controversy remains surrounding their accuracy and applicability within both legal and clinical contexts (Hart, 1998, 2003). Hart (1998) argued that the number of risk factors, incorporated within actuarial measures such as the SIR-R1 or the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993; Quinsey et al., 2006), is limited given the atheoretical and purely statistical selection process. Such selection processes may also ignore idiosyncratic risk factors considered to be “logical but of unknown validity” (p. 124). Similarly, the weighting of certain risk factors has been speculated to be inappropriate when applied within certain contexts (i.e., when assessing extremely violent offenders such as serial murderers; Hart, 1998; Litwack, 2001). For example, Hart (1998) noted that within the
VRAG, having a female victim and increased severity of victim injury are considered to be negatively related to violent recidivism risk. In assessing a serial sexual homicide offender (≥ 75 rapes and ≥ 3 serial sexual homicides), Hart found the offender posed "only an average or moderate risk for violent recidivism" (p. 125) according to his VRAG score. Following this counter-intuitive result, Hart noted that:

This is due in part to the fact that, all else being equal, killing women on the VRAG makes one a lower risk for violence than does, for example, threatening men; and because the VRAG fails to consider homicidal ideation or intent (e.g., sexual sadism). (p. 125)

Such as in the case of the serial sexual homicide offender who will likely never be released, assessing offenders who are considered unrepresentative of the development sample may result in what is considered an erroneous conclusion concerning their risk for future violence (Hart, 1998; Litwack, 2001). Hanson (2009) has argued that situations will undoubtedly arise that lead to the unavoidable, yet necessary, clinical interpretation of actuarial results as simply "scoring an actuarial risk tool is not a risk assessment" and that "[e]valuators will always need to make a separate judgment as to whether the risk scale score fairly represents the risk posed by the individual being assessed" (p. 174).

Other concerns raised relate to there being no clear discrimination among the types of violent behaviours the offender is at risk to commit (e.g., assault vs. murder; Jackson & Guyton, 2008; Litwack). Proponents of the actuarial risk assessment process (e.g., Hilton et al., 2010; Quinsey et al., 2006) argue, however, that measures such as the VRAG were not only designed to predict the nature and seriousness of violence (e.g., assault with a weapon, homicide, etc.), but have been found to be empirically related to such outcomes
ASSESSING RISK FOR INTIMATE PARTNER VIOLENCE

(e.g., Ho, Thomson, & Darjee, 2009). More recently, there has been considerable debate surrounding the applicability to individual cases the absolute probability estimates for recidivism derived from group data (for discussion see Hanson & Howard, 2010; Hart, Michie, & Cooke, 2007; Scurich & John, 2011). However, according to Hilton and colleagues (2010), “[w]hereas the absolute probabilities of recidivism reported for any sample may vary as a function of follow-up time and other variables, the rank order of perpetrators with respect to risk is more stable and generalizable” (p. 31). This assertion is somewhat questionable as research using probability bins developed for the VRAG have found inconsistencies with respect to the rank ordering of offenders relative to their estimated recidivism rates (Mills, Jones, & Kroner, 2005; Loza, Villeneuve, & Loza-Fanous, 2002).

While purely static actuarial risk measures may be of assistance in determining an offender’s level of risk, and hence, their level of treatment intensity (Harris & Rice, 2003), they remain uninformative with respect to the identification of treatment targets given their unchangeable nature (Andrews & Bonta, 2010; Hart, 1998; Litwack, 2001). Indeed, it is through the continued use of static risk factors that the assessment of violence risk remains bound to the offender’s past and does not reflect their current level of functioning or account for changes which may have resulted from treatment or life circumstances (Beggs & Grace, 2010; Douglas & Skeem, 2005; Hart, 1998; Yang, Wong, & Coid, 2010). Douglas and Skeem (2005) view risk assessments that strictly utilize static risk factors to be an indicator of an offender’s risk status, or simply, an indication of whether or not the offender poses a high risk for recidivism (i.e., purely predictive; Hilton et al., 2010). While informative in the assignment of treatment intensity, assessing
risk status remains uninformative with respect to the selection of treatment targets and monitoring of offenders over time (Douglas & Skeem, 2005).

In an attempt to enrich the risk assessment process by providing increased flexibility and accounting for offender change yet retaining a systematic and standardized method for assessing risk (Hart, 1998; Jackson & Guyton, 2008), a number of researchers began to develop risk assessment measures comprised of professional guidelines which were grounded in theory, empirical knowledge, and clinical practice (e.g., Kropp et al., 1999). Falling at the centre of the continuum described by Skeem and Monahan (2011), the process of utilizing professional guidelines, known as structured professional judgment (SPJ), provided an outline for the identification and measurement of risk factors yet allowed for flexibility within the risk assessment process. Flexibility was now afforded by the ability of the mental health professional to not only incorporate idiosyncratic risk factors considered pertinent to the current clinical context, but to use such factors as justification for overriding the level of assessed risk (Hart, 1998). However, this use of clinical override has been likened to introducing “noise” into the risk assessment process (Mills et al., 2011, p. 139; Quinsey et al., 2006) as a result of the detrimental impact it has on predictive accuracy (Hanson & Morton-Bourgon, 2009; Harris & Rice, 2003). Moreover, in contrast to actuarial measures, the estimated level of risk posed by the offender was no longer determined by probability estimates or percentage likelihoods; rather, mental health professionals were now required to make a descriptive determination (i.e., low, moderate, or high risk; Mills et al., 2011). While such a determination is solely based on the clinical opinion of the mental health professional, structured guidelines have been developed (Kropp et al., 1999).
Historically, items within SPJ guidelines have been selected based on a review of the empirical literature in addition to consultation with mental health professionals (Kropp et al., 1999; Webster, Eaves, Douglas, & Winstrup, 1995); this allows for greater generalizability across varying offender samples (Helmus & Bourgon, 2011). Apart from the pre-existing static risk factors, SPJ guidelines also incorporated what are considered to be dynamic risk factors known as criminogenic needs that, when targeted by treatment (e.g., antisocial attitudes), would result in a reduction in the likelihood that an offender would engage in future violent behaviour (Andrews & Bonta, 2010). Further subdividing dynamic risk factors, Hanson and Harris (2000) defined two distinct categories: stable dynamic and acute dynamic risk factors. Briefly, stable dynamic risk factors can change over long periods of time (within months or years) and include such factors as alcoholism, whereas acute dynamic risk factors change rapidly (within days, hours, or minutes) and are hypothesized to be related to the imminence or timing of recidivism in contrast to the long-term risk potential posed by the offender. Victim access is a risk factor that is considered to fall within the category of acute dynamic risk factors. Researchers now argue that there is evidence to such suggest that “risk ebbs and flows over time within each individual” (Douglas & Skeem, 2005, p. 348; Olver & Wong, 2011) and that by accounting for this variability in violence risk, through combining static and dynamic risk factors and conducting multiple time-point assessments, an offender’s risk state may be determined. Douglas and Skeem (2005) have argued that an offender’s risk state should be the focal point of the violence risk assessment process in contrast to assessing their risk status. Of note is that violence risk assessment measures developed within an SPJ framework are not the only measures capable of assessing risk
state as actuarial measures have been developed which incorporate both static and
dynamic risk factors that allow for "the remeasurement of ... dynamic risk factors [that]
can potentially alter the actuarial estimate of an individual's risk" (referred to as
dynamic-actuarial risk assessment; Mills et al., 2011, p. 22; Olver & Wong, 2011). To
date, however, there exists limited empirical research pertaining to the reliable
measurement and predictive utility of offender change (although see Beggs & Grace,
2010; Brown, Amond, & Zamble, 2009; Olver & Wong, 2011) as studies utilizing
measures with purportedly dynamic risk factors have predominantly consisted of single­
time point assessments, which, accordingly causes these measures to lose their dynamic
status thus becoming "static" in nature (Harris & Rice, 2003, p. 204). Regardless, SPJ
guidelines and actuarial risk assessment measures are currently the most common
methods used for assessing risk among offenders (Archer, Buffington-Vollum, Stredny,
& Handel, 2006).

Psychopathy and Violence Risk Assessment

The personality construct of psychopathy has frequently been incorporated into
the violence risk assessment process (Gendreau, Goggin, & Smith, 2002; Hilton et al.,
2010; Quinsey et al., 2006; Webster et al., 1995). Derived from Cleckley's (1976)
classical depiction of the prototypical psychopath, the Hare Psychopathy Checklist (PCL;
Hare, 1980) along with its revision (the Hare Psychopathy Checklist-Revised [PCL-R;
Hare, 1991, 2003]), derivative (the Hare Psychopathy Checklist: Screening Version
[PCL:SV; Hart, Cox, & Hare, 1995]) and downward extension (the Hare Psychopathy
Checklist:Youth Version [PCL:YV; Forth, Kosson, & Hare, 2003]) have become the
prominent measures for the assessment and diagnosis of psychopathic personality (Skeem
While earlier factor analyses conducted using the PCL/PCL-R found two underlying factors, recent analyses have revealed a hierarchical two-factor four-facet model of psychopathy (see Hare, 2003); however, this proposed factor structure has not been met with uniform support (see Cooke & Michie, 2001; Skeem & Cooke, 2010). Briefly, Factor 1, which encompasses Facet 1 (Interpersonal) and Facet 2 (Affective), consists of interpersonal and affective traits which reflect a “selfish, callous, and remorseless use of others,” whereas Factor 2, which encompasses Facet 3 (Lifestyle) and Facet 4 (Antisocial), consists of items tapping an unstable and socially deviant lifestyle and prior criminal behavior which are reflective of a “chronically unstable and antisocial lifestyle” or “social deviance” (Hare, 2003, p. 79).

Despite these diagnostic tools being designed for the strict purpose of assessing psychopathic personality (Hemphill & Hare, 2004), measures such as the PCL-R have been incorporated as items into violence risk assessment measures (e.g., VRAG and the HCR-20 [Webster et al., 1995]) or have become widely used by clinicians as stand-alone risk assessment measures (Archer et al., 2006; Viljoen et al., 2010). Andrews, Bonta, and colleagues (Andrews & Bonta, 2010; Andrews, Bonta, & Wormith, 2006) have argued that the risk assessment field’s increased interest in the PCL measures has resulted from the provision of a systematic and reliable process for measuring antisocial personality (Factor 1; hereafter referred to as F₁) and prior antisocial behavior (Factor 2; hereafter referred to as F₂); two of the most important areas for consideration in assessing risk for recidivism. Hart (1998), in reviewing the existing literature on psychopathy and violence, declared that “psychopathy is such a robust and important risk factor for violence that failure to consider it may constitute professional negligence” (p. 133).
Over time there have been a number of issues and concerns identified within the empirical literature pertaining to areas such as the misuse and pejorative nature of the psychopathic label/diagnosis (Hare, 1998), the designation of an appropriate diagnostic cut-off score (Salekin, Rogers, & Sewell, 1996), and poor inter-rater reliability among court- and defense-appointed forensic experts (Boccaccini et al., 2011). Given the controversy surrounding the use of PCL measures in the assessment of violence risk, a number of meta-analytic reviews have been conducted regarding its efficacy in predicting violent recidivism. One of the earliest meta-analytic reviews was conducted by Salekin and colleagues (1996) in which they found the PCL/PCL-R to be moderately predictive of violent behaviour (unweighted mean Cohen's $d = 0.79$, range = 0.42 to 1.92) and general recidivism (unweighted mean Cohen's $d = 0.55$, range = 0.24 to 0.93). Similarly, a meta-analysis conducted by Gendreau, Little, and Goggin (1996) found the PCL to be significantly predictive of recidivism ($k = 9, N = 1,040, Z^+ = .29$). Following their interpretation of the results, Salekin and colleagues stated that “[d]espite its limitations, the PCL-R appears to be unparalleled as a measure for making risk assessments with white male inmates” (p. 211). This controversial statement was later challenged in a meta-analysis conducted by Gendreau and colleagues (2002) who found the PCL/PCL-R to be a significantly poorer predictor of general and violent recidivism in comparison to the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995); however, this assertion was soon contested by Hemphill and Hare (2004). In contrast, Walters (2003) meta-analytically examined the predictive validity of $F_1$ and $F_2$ of the psychopathy checklists (i.e., PCL, PCL-R, PCL:SV, and PCL:YV) among prospective prediction studies and found that $F_2$ significantly outperformed $F_1$ in predicting violent recidivism.
(for $F_1$, $k = 27$, $r_w = .18$, 95% CI$_{rw}$ [.15, .20]; for $F_2$, $k = 27$, $r_w = .26$, 95% CI$_{rw}$ [.24, .29]). Significant differences were evidenced by the non-overlapping confidence intervals between the weighted effect sizes calculated for the two factors, in addition to a significant contrast analysis ($Z = 4.66$, Bonferroni corrected $p < .0125$, two-tailed).

Furthermore, Walters calculated the common language effect size indicator (CL; McGraw & Wong, 1992 – this indicates, within the present context, the probability that a randomly selected $F_2$ effect size will exceed a randomly selected $F_1$ effect size) and found that when randomly selecting effect sizes generated from $F_1$ and $F_2$, the effect sizes generated from $F_2$ would exceed those generated by $F_1$ approximately 74% of the time.

Similar effects were found within a recent meta-analysis conducted by Yang and colleagues (2010) who found $F_1$ to be a non-significant predictor of violent recidivism ($k = 13$, $N = 3,895$, $d = 0.22$, 95% CI$_d$[0.00, 0.45]), whereas $F_2$ was found to be a significant predictor ($k = 13$, $N = 3,995$, $d = 0.61$, 95% CI$_d$[0.38, 0.84]). In reflecting upon these results, they stated “these findings suggest that Factor 1 personality features, the core personality features of psychopathy, are not linked to violence. The predictive efficacy of the PCL-R appeared to be attributable almost entirely to Factor 2” (pp. 757-758).

Indeed, recent research conducted on the PCL and PCL-R indicates that the predictive accuracy of these measures appears to be related to a single facet (i.e., Facet 4 or Antisocial). Walters and colleagues (Walters & Heilbrun, 2010; Walters, Knight, Grann, & Dahle, 2008) utilizing a total of 7 non-overlapping forensic/correctional samples have found that Facet 4 added incrementally to the prediction of general and violent recidivism relative to the remaining three facets, yet Facets 1, 2, and 3 did not add
incrementally relative to Facet 4. This finding was most recently replicated by Gray and Mills (2011) who extended the follow-up period and recoded recidivism rates for one of the samples (i.e., Kroner et al., 2005) used within the analyses by Walters and colleagues (2008). In addition, Gray and Mills found that the predictive accuracy of the HCR-20 and VRAG was not dependent upon the PCL-R total score when incorporating the Antisocial Facet score in its place. This finding was consistent with other studies which had either removed or incorporated other measures in place of the PCL-R total score (e.g., Glover, Nicholson, Hemmati, Bernfeld, & Quinsey, 2002; Guy, Douglas, & Hendry, 2010; Harris, Rice, & Quinsey, 1994; Kroner & Mills, 2001). In summary, empirical research now suggests that it is not the core psychopathic personality features, encompassed within Facets 1, 2, and 3 of the PCL/PCL-R, that are predictive of violent recidivism, rather, it is the prior criminal history which is captured within the Antisocial facet.

Rufino, Boccaccini, and Guy (2011) found Facet 4 of the PCL-R to be rated as requiring the least amount of subjectivity (i.e., subjective judgment) to rate the items in comparison to Facets 1, 2, and 3. Moreover, they found that as levels of inter-rater reliability increased, subjectivity decreased. Given that increased inter-rater reliability is associated with increased predictive accuracy (Hanson & Morton-Bourgon, 2009), Rufino and colleagues have argued that the level of subjectivity required to score the items within Facets 1, 2, and 3 relative to Facet 4 may account for the low predictive accuracy. While this may be a plausible explanation given the poor inter-rater reliability among court- and defense-appointed forensic experts, it does not account for the high levels of inter-rater reliability found for the PCL-R within research (e.g., Kroner & Mills, 2001; ICC for PCL-R total score = .97) and applied contexts (e.g., Looman & Ismail, 2012; ICCs = .90,
Overall, psychopathic personality as assessed by the PCL measures remains a central feature of studies related to violent and antisocial behaviour and is an important consideration with respect to treatment responsivity among violent offenders (Olver & Wong, 2009).

**Risk Assessment for IPV**

As originally observed by Kropp (2004) and later by Bowen (2011), the field of IPV risk assessment has “lagged” in the past and continues to “lag” in comparison to other areas such as general violence and sexual violence with the majority of IPV risk assessment measures being developed only within the past 12 years (see Hanson, Helmus, & Bourgon, 2007; Kropp, 2008). A possible explanation for this lag in the development and validation of IPV risk measures may be related to researchers questioning the necessity of IPV risk assessment measures given that the risk factors for violent recidivism tend to be related to the risk factors for domestic violence (Hanson et al., 2007; Hilton & Harris, 2005; Hilton, Harris, & Rice, 2010; Grann & Wedin, 2002). This issue of multicollinearity or redundancy, however, comes as no surprise as it has been previously observed between risk factors related to general and violent recidivism (see Kroner, Mills, & Reddon, 2005) across varying offender populations such as sex offenders (Hanson & Morton-Bourgon, 2009), mentally disordered offenders (Bonta, Law, & Hanson, 1998), and IPV offenders (Hanson & Wallace-Capretta, 2004; Hilton, Harris, & Rice, 2001). One risk assessment measure which was designed to predict violent recidivism which has been found to be moderately predictive of IPV is the VRAG. Within their meta-analysis, Hanson and colleagues (2007) found that the VRAG displayed the largest association with IPV recidivism ($k = 2, N = 736, d = 0.65, 95\% \text{ CI}_d$).
[0.49, 0.80]) when compared to other risk assessment measures designed specifically for predicting IPV. Although some of the items incorporated within the VRAG such as failure on prior conditional release and substance abuse score are considered predictive of future IPV (and have subsequently been incorporated into the ODARA and DVRAG – see below), other items such as having a female victim and having ever been married were unable to distinguish between IPV recidivists and non-recidivists and were considered inappropriate within the context of IPV (Hilton et al., 2010).

Despite this, Kropp (2008) noted that during the past decade there has been a "proliferation" of risk assessment measures designed for IPV offenders. A caveat, however, is that "[m]any of these measures were developed by local correctional, police, or victim agencies frontline agencies" with the developers having "not reported normative, reliability, and validity data" (Kropp, 2008, p. 208). Thus, Kropp (2008), considered only the ODARA, SARA, Domestic Violence Screening Inventory (DVSI; Williams & Houghton, 2004), and the Danger Assessment (DA; Campbell, 1995) to "hold the most promise" (p. 208). Indeed, within an extensive review of the literature on IPV risk assessment, only the development, reliability, and predictive validity of the DA, DVSI, ODARA (and DVRAG), SARA, and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp & Hart, 2004; a derivative of the SARA designed for use by frontline police personnel) was reviewed as these were the only measures with studies meeting the following criteria: used an adult sample, was published in English language, and published in a peer reviewed journal (Bowen, 2011). Therefore, to date, the most widely researched and validated IPV risk assessment measures are the DA, DVSI, ODARA (and subsequently the DVRAG), and SARA.
The SARA, one of the earliest developments to address the need for a specialized risk assessment measure for IPV offenders, is arguably the most widely used measure of IPV risk and is currently used to assess all Canadian federal offenders suspected of or who have a documented history of IPV. Embedded within the SPJ framework, the SARA was developed following a review of the empirical literature on IPV recidivism and consultation with experts in the area of IPV risk assessment. The SARA consists of 20 static and dynamic risk factors related to IPV recidivism which are embedded within four main sections: Criminal History, Psychosocial Adjustment, Spousal Assault History, and Alleged/Most Recent Offense. According to Kropp and Gibas (2009) a risk factor was required to meet one or all of the following three criteria to be selected for inclusion in the SARA:

1. There was compelling evidence in the empirical literature that the risk factor discriminated those who were violent toward spouses from those who were not.

2. There was evidence that the risk factor was associated with recidivistic spousal violence.

3. The risk factor was included in other well-organized professional guidelines for spousal violence risk assessment. (p. 228)

Also provided within the SARA guidelines is an additional section titled Other Considerations which allows for the mental health professional to note important risk factors not incorporated within the SARA (e.g., history of pedophilia, history of sexual sadism, history of stalking, history of disfiguring, torturing, or maiming intimate partners, etc.). Although items are rated on a 3-point scale, the overall risk designation (or
summary risk judgment of low, moderate, or high) is generally determined by the opinion of the mental health professional with no specific rules for combining or weighting risk factors. A total score for the SARA can be derived by simply summing the 20 items which can then be used in conjunction with descriptive cutoffs available within the manual. Moreover, depending on the clinical context, the SARA allows for an item to be coded as critical and, hence, makes items “sufficient on their own to compel the evaluator to conclude that the individual poses an imminent risk of harm” (Kropp & Gibas, 2009, p. 232). Research conducted with the SARA has found acceptable inter-rater reliability among the total scores and summary risk judgments, but low inter-rater reliability among the critical items (Helmus & Bourgon, 2011). Furthermore, the SARA has shown good criterion-related and concurrent validity (Helmus & Bourgon, 2011; Kropp & Gibas, 2009; Kropp et al., 1999). See below for a discussion regarding the predictive validity of the SARA.

The 12-item DVSI was originally designed as a brief screening device completed by probation services to inform court dispositions and to recommend whether the administration of the SARA is required. The DVSI has since undergone revision and is now referred to as the Domestic Violence Screening Inventory-Revised (DVSI-R; Williams, 2011; Williams & Grant, 2006). However, revisions made to the 11 items now within the DVSI-R were related to broadening the focus of the items to incorporate intimate violence towards family members such that violence within parent, child, sibling, intimate partner, or “other family or family-like relationships” was accounted for as opposed to strictly violence towards an intimate partner (Williams & Grant, 2006, p. 401). The DVSI-R has since been cross-validated using an American sample comprised
of both male and female IPV offenders \((N = 3,569)\) and has been found to be moderately to highly predictive of family violence recidivism (Bowen, 2011; Williams, 2011). In contrast to the other four IPV risk measures, the DA (available online at www.dangerassessment.com/DATools.aspx) is scored entirely by the victim and is comprised of two sections. The first section consists of a calendar spanning across the past year in which the victim is instructed to record the dates in which she was abused by her current or ex-partner. In addition, the victim is to rate the severity of the incidents based on a scale provided (ranges from 1 = slapping, pushing; no injuries and/or lasting pain, to 5 = use of a weapon; wounds from weapon). For the second section, the victim completes 20 questions in a yes/no format relating to prior incidents of IPV (e.g., Does he threaten to kill you?) or characteristics of the perpetrator (e.g., Does he own a gun?). Hanson and colleagues (2007) found the DA to be a modest predictor of IPV recidivism \((k = 4, N = 1,585, d. = 0.41, 95\% \text{ CI}_d [0.31, 0.52])\). Notwithstanding the reliability and validity of the DVSI-R and DA, they are excluded from further discussion within the present study given the focus of the DVSI-R on family violence as opposed to strictly IPV in addition to the DA solely being scored by the victim.

**Development of the ODARA and DVRAG**

Despite the existence of several IPV risk assessment measures such as the SARA (see Kropp, 2004, 2008), at the time of development there were no pre-existing actuarial measures designed to assess IPV risk (Hilton et al., 2010). In their rationale for developing the Ontario Domestic Assault Risk Assessment (ODARA), Hilton and colleagues (2004) acknowledged that while the VRAG had been found to be predictive of IPV recidivism, law enforcement personnel such as police officers would be unable to
score offenders on the VRAG given the extensive information required (e.g., PCL-R score, psychological characteristics, and criminal history). Therefore, in developing the ODARA, item development was based on information contained within the reports completed by the Ontario Provincial Police (OPP) for the purpose of scoring the Domestic Violence Supplementary Report (DVSR; Ministry of the Solicitor General, 2000) an IPV risk assessment measure which is mandatory among all Ontario police services (Heaton, 2010). To score the DVSR, OPP officers are required to conduct interviews with victims of an IPV incident and conduct a thorough review of the perpetrators’ criminal record (all information was stored within the Ontario Municipal Provincial Police Automated Cooperative [OMPPAC] system). Utilizing information from both the OMPPAC system and the Canadian Police Information Centre (CPIC; which contains information pertaining to the criminal charges, arrests, and convictions for all offenders derived from the Fingerprint Service [FPS] records of the Royal Canadian Mounted Police), Hilton and colleagues identified a sample of 689 offenders (589 designated for the construction phase and an additional 100 for the cross-validation process) in which there was the presence of “a victim report or police evidence of forceful physical contact by a man against his current or former wife or common-law wife” (p. 269). All events of domestic violence occurring prior to December 31, 1996 were recorded. Domestic violence recidivism was then coded, regardless of whether a formal charge was entered, over a mean follow-up period of 4.79 years (from December 31, 1996 to the end of 2001) of which 30% ($n = 175$) of the 589 offenders within the construction sample re-offended. Of note is that approximately 95% of the recidivists offended against the victim of their index offence.
Using the same statistical procedure employed to develop the VRAG such as setwise (i.e., stepwise analyses that are applied to multiple sets of items) and stepwise selection analyses, Hilton and colleagues (2004) identified a total of 13 items based on their unique statistical contribution to the prediction of domestic violence recidivism. A weakness of this method is that such analyses capitalize on variance specific to the sample, thus to reduce shrinkage (i.e., reduced predictive validity) during the cross-validation process the authors utilized bootstrapping procedures which derived nine randomly selected subsamples of 359 cases, with replacement, from the total construction sample. Once drawn, each subsample was subjected to a forward conditional logistic regression analysis. Retaining only those variables which were selected in at least five of the nine regression analyses, the forward conditional logistic regression procedure was again repeated for all of the 9 subsamples in addition to the total construction sample \((n = 589)\). Variables excluded from the prediction models computed for the 9 subsamples and total construction sample were dropped from any further analyses. In total, 54 setwise and stepwise selection analyses were conducted (for further details see Hilton et al., 2004). The resulting 13-item measure became the ODARA (see Table 2). Inter-rater reliability of the continuous and categorical ODARA items yielded correlation coefficients of .80 and above and kappa coefficients of .70 and above, respectively (note that all items were eventually dichotomized). Following the dichotomization of the items, the range of total scores for the ODARA was restricted between 0 and 13. Moreover, to ensure that the presence of IPV recidivism data did not unduly influence the scoring of the predictor variables, inter-rater reliability was assessed on a subsample of 24 cases yielding high correlations between ODARA scores (intra-class correlation coefficient...
Table 2

*Items from the Ontario Domestic Assault Risk Assessment (ODARA) and Domestic Violence Risk Appraisal Guide (DVRAG)*

<table>
<thead>
<tr>
<th>Item</th>
<th>ODARA</th>
<th>DVRAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior domestic incident</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Prior nondomestic incident</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Prior custodial sentence of 30 days or more</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Failure on prior conditional release</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Threat to harm or kill at the index assault</td>
<td>0, 1</td>
<td>0</td>
</tr>
<tr>
<td>Confinement of partner at index assault</td>
<td>0, 1</td>
<td>0</td>
</tr>
<tr>
<td>Victim concern</td>
<td>0, 1</td>
<td>0</td>
</tr>
<tr>
<td>More than one child</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Victim’s biological children from a previous partner</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Violence against others</td>
<td>0, 1</td>
<td>0</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>0, 1</td>
<td>-2</td>
</tr>
<tr>
<td>Assault on victim when pregnant</td>
<td>0, 1</td>
<td>0</td>
</tr>
<tr>
<td>Barriers to victim support</td>
<td>0, 1</td>
<td>-1</td>
</tr>
<tr>
<td>Psychopathy Checklist-Revised score</td>
<td>-</td>
<td>-1</td>
</tr>
</tbody>
</table>

[ICC] = .90 and coded recidivism (ICC = .91). The calculated standard error of measurement (SEM) was less than 1 point (i.e., ± .94). Moreover, to ensure reliability in scoring the ODARA among law enforcement personnel, two police officers (with brief training and who were independent of the development of the ODARA) scored 10 cases each in which the ICC was .95, thus providing evidence that the ODARA would be a reliable measure. A total of seven probability bins were derived for the possible range of ODARA scores (recidivism rate = .05, .10, .20, .27, .41, .59, and .70 for categories 1 to 7, respectively). Validity of the ODARA was demonstrated by significant associations with other validated measures of IPV risk (e.g., \( r = .60 \) and .53 for the SARA and DVSR, respectively). Similarly, ODARA total scores were significantly associated with other
variables such as increased severity of violence, number of incidents with severe violence, and time at risk (see below for further discussion regarding predictive validity).

Notwithstanding the accuracy of the ODARA in predicting IPV recidivism, Hilton and colleagues revised the measure in 2008, arguing that not accounting for more indepth clinical information, such as extensive information pertaining to an offender’s history of criminal and antisocial behaviour, prior relationships, and personality characteristics (which is generally unavailable to frontline police personnel), “could lead to suboptimal prediction” (p. 151). Using 303 cases with extensive file information from the ODARA construction sample with an additional sample of 346 cases meeting inclusion criteria, a number of risk assessment measures (e.g., VRAG, SARA, PCL-R, etc.) along with the ODARA were entered into a series of logistic and linear regressions with various measures of recidivism as the outcome. With respect to dichotomous recidivism, none of the measures provided any independent improvement or added any incremental validity to the ODARA. For the linear regression analyses, however, the PCL-R was found to significantly improve the performance of the ODARA in predicting various recidivism outcomes (i.e., number of recidivistic incidents, number of incidents with severe violence, and for total recidivism injury). With the exception of one continuous outcome (pertaining to the Cormier-Lang Criminal History score; Quinsey et al., 2006), the PCL-R was considered to be the best predictor in comparison to the other risk measures and was the most consistent in improving upon the predictive validity of the ODARA. Using the simple summation technique previously used by Nuffield (1982), statistical weights were developed for each item based on the deviation from the recidivism base-rate (see Table 2). The mean score of the combined sample ($n = 649$) was
2.88 with scores ranging from -10 to +37. Associations between the DVRAG and the various forms of recidivism were considered to be a significant improvement over those found for the ODARA, and, as seen previously with the ODARA, inter-rater reliability was high among independent raters (ICC = .90).

Despite its performance within the DVRAG sample, the incorporation of the PCL-R score appears counter-intuitive considering that psychopathy has long been speculated to be unrelated to IPV as a "fundamental" aspect of psychopathic personality is the absence of any meaningful interpersonal attachment with others (Dutton & Kropp, 2000). Similarly, psychopathic personality as measured by the PCL-R has been found to be significantly lower in IPV offenders compared to non-IPV offenders (Hilton et al., 2001) and a comparison between recidivistic and non-recidivistic IPV offenders revealed non-significant differences for the PCL:SV total and factor scores (Kropp et al., 1999). A recent study by Storey, Hart, Meloy, and Reavis (2009) found psychopathy to be almost non-existent among a sample of 61 adult stalkers with only a single participant meeting the criteria for diagnosis (i.e., score of ≥ 18; scores ranged from 2 to 18, M = 8.13, SD = 3.50). Indeed, Mills and colleagues (2011) noted that the relatively low mean PCL-R scores within the DVRAG development sample (M = 8.4, SD = 6.7) indicates "that approximately 80% of the spouse abuse sample had scores below that of the average violent criminal" (p. 107). Rettenberger and Eher (2012), in cross-validating the ODARA and DVRAG, found that the PCL-R neither added any significant incremental validity to the ODARA nor significantly predicted violent or IPV recidivism. In contrast, however, at least two studies have found the PCL-R to be predictive of IPV recidivism (Grann & Wedin, 2002; Hilton et al., 2008; see below), while another found that those who
committed IPV recidivism following treatment “were six times as likely to have PCL:SV
scores above 12” (Dutton & Kropp, 2000, p. 177).

A Meta-Analytic Investigation into the Accuracy of Predicting IPV

Within the empirical literature, the area under the curve (AUC) of the receiver
operating characteristic (ROC) has become the preferred method by which the predictive
accuracy of violence risk assessment is measured (Mossman, 1994; Rice & Harris, 1995;
Rice & Harris, 2005; Swets, Dawes, & Monahan, 2000). Calculated from observed
frequencies (e.g., two-by-two contingency tables), the ROC curve represents the
proportion of the true-positive rates (i.e., hit rate or sensitivity) as a function of the false-
positive rates (i.e., 1-false alarm rate or specificity) along a continuum of various scale
cutoff scores (Rice & Harris, 1995; Swets et al., 2000). The accuracy of a risk assessment
measure, however, is represented by the calculated proportion of the total area under the
ROC curve (i.e., AUC; Hanley & McNeil, 1982). The AUC value measures the
probability that a randomly selected recidivist will score higher on a risk assessment
measure than a randomly selected non-recidivist (Hanley & McNeil, 1982; Rice &
Harris, 1995). AUC values can range between 0 and 1, with .5 representing accuracy that
is no better than chance (Swets et al., 2000). In comparison to other effect sizes such as
Pearson’s $r$ or Cohen’s $d$, AUC values are robust against violations of normality and
fluctuations in base rate and selection ratios, and are easily interpreted (Mossman, 1994;
Rice & Harris, 1995; Rice & Harris, 2005). Rice and Harris (2005) determined that AUC
values of .556, .639, and .714 are reflective of small, moderate, and large effect sizes,
respectively.
To date, there has been a limited amount of empirical research conducted on the accuracy of structured risk assessment measures in predicting IPV. In their meta-analysis on the accuracy of risk assessment measures in predicting IPV recidivism, Hanson and colleagues (2007) could only identify 18 independent studies that met their inclusion criteria. However, Helmus and Bourgon (2011) were able to update the meta-analysis by adding an additional four studies examining the predictive accuracy of the SARA total score. For the purpose of the current literature review, previous meta-analytic studies were updated to determine the current effect sizes associated with the study measures. AUC values and base rates were obtained from all studies identified by Helmus and Bourgon (2011) and Kropp and Gibas (2009) for the SARA (Andrés-Pueyo, López, & Álvarez, 2008; Gibas, Kropp, & Hart, 2008; Grann & Wedin, 2002; Heckert & Gondolf, 2004; Hilton, et al., 2008; Kropp & Hart, 2000; Williams & Houghton, 2004; Wong & Hisashima, 2008), and all studies identified by the Waypoint Centre for Mental Health (2011) for the ODARA and DVRAG (Buchanan, 2009; Hilton & Harris, 2009; Hilton, Harris, Popham, & Lang, 2010; Hilton et al., 2008; Stewart & Henning, 2010; Trinh, 2010). Two additional studies recently published by Belfrage and colleagues (2012) and Rettenberger and Eher (2012) utilizing the SARA and the ODARA/DVRAG, respectively, were also incorporated into the meta-analysis. Additionally, three studies were identified which examined the predictive accuracy of the PCL-R with IPV recidivism (Grann & Wedin, 2002; Hilton et al., 2008; Rettenberger & Eher, 2012; the first two were identified by Hanson et al., 2007), whereas only a single study was located which examined the SIR-R1 in predicting IPV recidivism (Gibas et al., 2008). Although Hanson and colleagues excluded results derived from the development samples of the
ODARA and DVRAG, these results were included within the current analysis due to the limited number of studies identified for inclusion.

Point-biserial correlations were converted into AUC values using formulas provided by Rice and Harris (2005), whereas the Wilcoxon standard error, SE($W$), was calculated for each AUC value using the formula described by Hanley and McNeil (1982). Statistical techniques described by McClish (1992) were used to calculate the mean weighted effect sizes ($\text{AUC}_w$) and to test the equality of the AUC values across each measure. Consistent with procedures first outlined by Hedges and Olkin (1985), each AUC value was first weighted by the inverse of its variance: $W_i = 1/[\text{SE}(W)]^2$. Formula 1 from McClish (p. 274) was then used to calculate the mean weighted effect size: $\text{AUC}_w = \frac{\sum (W_i \times \text{AUC}_i)}{\sum (W_i)}$.

To test the significance and magnitude, a z-test and the corresponding 95% confidence intervals were calculated for each mean weighted AUC value using its respective weighted standard error (Formula 2 from McClish, 1992, p. 1): $\text{SE}(\text{AUC}_w) = \sqrt{1/\sum (W_i)}$.

Finally, to test the equality (or homogeneity) of the AUC values across each measure, a chi-square ($\chi^2$) statistic (hereafter referred to as the $Q$ statistic; see Hedges & Olkin, 1985; Rosenthal, 1991) was calculated for each mean weighted effect size using Formula 4 from McClish: $\sum W_i \times (\text{AUC}_i - \text{AUC}_w)^2$. The significance of the $Q$ statistic is then determined as a $\chi^2$ distribution with $k - 1$ degrees of freedom. A significant $Q$ statistic is an indication that the AUC values are heterogeneous and do not support the null hypothesis that all of the values originate from the same population (McClish, 1992).

---

2 It was noted that McClish (1992) uses the weighted variance to calculate the 95% confidence intervals for the mean weighted AUC value; however, in keeping with Hanley and McNeil (1982) the square root of the weighted variance (i.e., the weighted standard error) was used to calculate the 95% confidence intervals within the current analysis.
All calculations were performed using SPSS syntax (see Appendix A) originally developed by Andy Field (available at statisticshell.com) but revised by the present author to incorporate formulae described by McClish (1992). Meta-analytic results for the five risk assessment measures in predicting IPV are reported in Table 3. The average base rates for IPV recidivism were as follows: DVRAG = 34%, k = 4; ODARA = 33%, k = 7; PCL-R = 28%, k = 3; SARA = 34%, k = 9; and the SIR-R1 = 28%, k = 1.

Table 3

<table>
<thead>
<tr>
<th>Measure</th>
<th>n (k)</th>
<th>AUC&lt;sub&gt;w&lt;/sub&gt;</th>
<th>95% CI&lt;sub&gt;lw&lt;/sub&gt;</th>
<th>z</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVRAG</td>
<td>779 (4)</td>
<td>.709</td>
<td>[.671, .748]</td>
<td>10.73*</td>
<td>0.73</td>
</tr>
<tr>
<td>ODARA</td>
<td>1,423 (7)</td>
<td>.678</td>
<td>[.648, .708]</td>
<td>11.64*</td>
<td>6.92</td>
</tr>
<tr>
<td>PCL-R</td>
<td>802 (3)</td>
<td>.658</td>
<td>[.618, .698]</td>
<td>7.81*</td>
<td>1.67</td>
</tr>
<tr>
<td>SARA</td>
<td>2,602 (9)</td>
<td>.633</td>
<td>[.609, .657]</td>
<td>10.88*</td>
<td>14.65</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>108 (1)</td>
<td>.464</td>
<td>[.344, .584]</td>
<td>0.59</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note. n = total sample size; k = number of effect sizes; AUC<sub>w</sub> = mean weighted effect size; CI<sub>lw</sub> = 95% confidence interval of AUC<sub>w</sub>; z = significance test of AUC<sub>w</sub>; Q = test of homogeneity. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1.
*p < .001.

As displayed within Table 3, the DVRAG yielded the highest predictive accuracy in comparison to the remaining four risk assessment measures (AUC<sub>w</sub> = .709, 95% CI<sub>lw</sub> [.671, .748]). While only based on four effect sizes (including one which originated from the development sample; Hilton et al., 2008), the mean weighted AUC for the DVRAG is considered to be a large effect based on the values provided by Rice and Harris (2005). Similarly there was no heterogeneity among the AUC values identified for the DVRAG.
Indeed, none of the remaining analyses in which the $Q$ statistic could be calculated achieved significance, therefore, the AUC values aggregated for each measure were considered homogeneous. Moderate predictive accuracy was found for the ODARA ($AUC_w = .678, 95\% CI_w [.648, .708]$) and PCL-R ($AUC_w = .658, 95\% CI_w [.618, .698]$), whereas the SARA ($AUC_w = .633, 95\% CI_w [.609, .657]$) fell just below the threshold (i.e., .639). However, a caveat to this result is that prior meta-analytic research has revealed that the predictive accuracy for the SARA is improved when the clinical judgments are examined in comparison to the calculated total scores (Helmus & Bourgon, 2011). Interestingly, despite the SIR-R1 scale exhibiting moderate predictive accuracy for violent recidivism (e.g., Campbell et al., 2009), when predicting IPV recidivism the SIR-R1 scale exhibited worse than chance accuracy ($AUC_w = .464, 95\% CI_w [.344, .584]$). However, caution must be used when interpreting this finding as the effect size was generated from a single study. While the poor predictive accuracy exhibited by the SIR-R1 can only be considered tentative at this time, a lack of association has been previously found between the SIR-R1 and IPV risk factors such as those found on the SARA (Kropp et al., 1999). Similarly, while measures designed to assess risk for violence such as the VRAG are predictive of IPV recidivism, there is evidence within the empirical literature to suggest that risk assessment measures designed for predicting general recidivism (such as the SIR-R1) are not as accurate at predicting varied forms of violence such as sexual recidivism (Hanson & Morton-Bourgon, 2009) and IPV recidivism (Hilton, Harris, Popham, et al., 2010). Therefore, it could be argued that the two constructs (i.e., risk for general recidivism and risk for IPV recidivism) do not overlap.
Moderator Analysis

Excluding the SIR-R1 scale from any further analyses, the overall level of variability among the effect sizes was examined. To facilitate this, the individual effect sizes for the DVRAG, ODARA, PCL-R, and SARA were combined and re-examined (overall $Q = 36.62, df = 22, p = .026$); note that this form of analysis increases error as it ignores the independence of effect sizes within samples (see Hanson & Morton-Bourgon, 2009). A between-level $Q$ ($Q_\Delta$) was calculated for the combined analysis and revealed a significant moderating effect of risk assessment measure used ($Q_\Delta = 12.65, df = 3, p = .006$). Visual examination of the confidence intervals between the four measures indicated that the DVRAG was a significantly superior predictor of IPV recidivism in comparison to the SARA ($p \leq .01$; see Cumming & Finch, 2005). However, to further explore whether there were any significant differences between the measures a between-level $Q$ was calculated for each possible pairwise comparison (see Table 4).

Table 4

<table>
<thead>
<tr>
<th>Pairwise comparison (k)</th>
<th>AUC\text{between}</th>
<th>$Q_\Delta$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVRAG versus ODARA (11)</td>
<td>.71</td>
<td>1.60</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>DVRAG versus PCL-R (7)</td>
<td>.83</td>
<td>3.32</td>
<td>1</td>
<td>.068</td>
</tr>
<tr>
<td>DVRAG versus SARA (13)</td>
<td>.90</td>
<td>11.00</td>
<td>1</td>
<td>.0009</td>
</tr>
<tr>
<td>ODARA versus PCL-R (10)</td>
<td>.64</td>
<td>0.61</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>ODARA versus SARA (16)</td>
<td>.72</td>
<td>5.28</td>
<td>1</td>
<td>.022</td>
</tr>
<tr>
<td>PCL-R versus SARA (12)</td>
<td>.56</td>
<td>1.14</td>
<td>1</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Note. $k =$ number of effect sizes per pairwise comparison; $Q_\Delta =$ between-level $Q$. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment. Bonferroni adjustment: $p < .05 = .008, p < .01 = .002, p < .001 = .0002$. 
As displayed within Table 4, significant differences in predictive accuracy were found between the DVRAG and the SARA, and between the ODARA and the SARA. Similarly, differences between the DVRAG and the PCL-R appeared to be approaching significance \( (p = .068) \). After applying the Bonferroni adjustment for the multiple pairwise comparisons, significant differences remained only between the DVRAG and the SARA. Similar to Walter's (2003) use of the CL effect size, AUC values were calculated for each pairwise comparison (represented by AUC\textsubscript{between} in Table 4) such that the AUC\textsubscript{between} values represent the probability that a randomly selected effect size from the first measure will be higher in value in comparison to a randomly selected effect size from the second measure. For example, comparisons between the effect sizes generated by the DVRAG with those generated by the SARA yielded an AUC\textsubscript{between} of .90; therefore, when randomly selecting effect sizes generated from the DVRAG and SARA, the effect sizes generated from the DVRAG would exceed those generated by the SARA approximately 90% of the time. Note that only the AUC\textsubscript{between} calculated for the DVRAG and SARA achieved significance \( (p = .025) \).

While these results suggest that the SARA is significantly less accurate at predicting IPV recidivism in comparison to the ODARA and DVRAG, Kropp and Gibas (2009) noted that three of the prediction studies utilizing the SARA did not administer it as outlined within the manual (i.e., Grann & Wedin, 2002; Heckert & Gondolf, 2004; Hilton et al., 2008). Although the significance test of the Q-statistic did not reveal significant variability among the effect sizes at the \( p < .05 \) level, the value did appear to be approaching significance \( (p = .066) \). Therefore, the percentage of overall variability among the effect sizes for the SARA was calculated and was revealed to be within the
low to moderate range ($R^2 = 45\%$; see Higgins, Thompson, Deeks, & Altman, 2003). To determine if administering the SARA as outlined within the manual had an effect on its predictive accuracy, two separate meta-analyses were run on the six studies which did follow manual instructions ($k = 6$, $\text{AUC}_w = .656$, 95% CI$_w$ [.622, .689]) versus the three that did not ($k = 3$, $\text{AUC}_w = .609$, 95% CI$_w$ [.574, .643]). The between-level $Q$ revealed that there was indeed a detrimental impact on the weighted AUC value when incorporating studies which did not follow the manual instructions; although, the level of significance was just slightly above the threshold of the $p = .05$ level ($Q_A = 3.65$, $df = 1$, $p = .056$). Regardless of the level of significance, excluding the three studies exhibiting poor methodology in relation to the SARA did provide an increase in the size of the weighted AUC value ($\text{AUC}_w = .656$; a moderate effect size) such that the predictive accuracy of the DVRAG was no longer significantly higher at the $p \leq .01$ level as evidenced by the overlapping confidence intervals. Although these results are based solely on 23 effect sizes, they do suggest that the DVRAG (in comparison to the PCL-R and SARA) is currently the most efficacious predictor of IPV recidivism.

**Study Purpose and Hypotheses**

The purpose of this study was to investigate the reliability and validity of two actuarial risk assessment measures designed for domestically violent offenders: the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004; Hilton et al., 2010) and the Domestic Violence Risk Appraisal Guide (DVRAG; Hilton et al., 2010; Hilton et al., 2008). To date, only a single study by Hilton, Harris, Popham, and colleagues (2010) has assessed the validity and reliability of the ODARA among a mixed sample of provincially and federally incarcerated offenders ($n = 150$) referred for
specialized IPV programming between 1995 and 2000. The mean ODARA score for the entire sample was found to be within the second highest risk category and was significantly higher in comparison to previously published norms. An unfortunate limitation of the study was the authors’ inability to assess the DVRAG due to limited PCL-R data. Therefore, the intent of this study was to address this limitation and subsequently build upon the pre-existing research with the ODARA and DVRAG.

Further, this study allowed for concurrent and predictive validity analyses between the ODARA, DVRAG, PCL-R, SARA, and the SIR-R1 in predicting recidivism (i.e., any violent recidivism, and more specifically IPV recidivism) among Canadian federal offenders. This has implications regarding the assessment of IPV offenders, particularly as it relates to the ODARA, DVRAG, and the SARA. The latter of which is utilized on a national level by the Correctional Service of Canada (CSC) with all offenders who commit violence against an intimate partner.

**Hypotheses**

Within the current study, the following hypotheses were made:

1. The ODARA and DVRAG would be reliable with respect to both their structural reliability (i.e., internal consistency, corrected item-total correlations) and inter-rater reliability.

2. The ODARA and DVRAG would have concurrent and discriminant validity as evidenced by the following: their moderate to large positive association with a pre-existing measure of IPV risk which has been validated among federal offenders (i.e., the SARA) and lower association with assessed risk for general recidivism (i.e., the SIR-R1).
The ODARA and DVRAG would have moderate to high predictive validity in predicting IPV recidivism (i.e., predictive validity).

**Power Analysis and Proposed Sample Size**

To investigate the predictive validity of the DVRAG and ODARA, this study employed a hybrid prospective-retrospective design with a subsample of offenders originating from a previously identified sample of 586 male federal offenders who met the criteria for referral to either the Moderate or High Intensity Family Violence Prevention Program (MIFVPP and HIFVPP, respectively; Connors, Mills, & Gray, 2012; Connors, Mills, & Gray, in press). According to the Psychology of Criminal Conduct (PCC; Andrews & Bonta, 2010) there exists a potential for treatment to have a confounding effect on the predictive accuracy of the risk assessment measures (via reductions in base rate). Indeed, prior empirical research has revealed that higher risk offenders tend to have more "room" to move with respect to their level of dynamic risk and, hence, may exhibit increased treatment change leading to reduced rates in violent recidivism (e.g., Andrews & Bonta, 2010; Olver & Wong, 2011); however, this has been contested with respect to IPV offenders (Hilton et al., 2010). Reviews of the literature have found that IPV treatment to date has little to no effect and that in some cases it has had effects in the opposite direction (i.e., increases IPV recidivism; Hilton & Harris, 2005; Hilton et al., 2010). Indeed, among two evaluations of the FVPP, one study conducted by Kropp and Lee (2004) found significant reductions at the \( p < .05 \) level in the prevalence rates for IPV recidivism between treated versus untreated (included treatment drop outs) HIFVPP participants (4% for treated and 14% for untreated, Odds

---

3 However, correlational analysis between the observed AUC values and base rates for the studies using the DVRAG, ODARA, PCL-R, and SARA \( (k = 23) \) reveals that the two are unrelated \( r = .04, Z = 0.64, p > .05 \).
Ratio \([OR = 4.50, p = .033]\), but not among the MIFVPP participants (4% for treated and 12% for untreated, \(OR = 3.25, p = 3.25\)). In contrast, a more recent evaluation of the two programs conducted in 2009 by CSC found non-significant hazard ratios (HR) for violent recidivism when comparing the intent-to-treat group which consisted of offenders “who were exposed to the targeted correctional program through varying degrees of program participation” (Evaluation Strategy section, para 4) to those who did not participate (i.e., comparison group) in the HIFVPP (HR = 1.67 and 1.00 for intent-to-treat and comparison groups, respectively). Although non-significant, this difference would indicate that the offenders who participated to some degree in the HIFVPP actually had increased rates of violent recidivism and were 67% more likely to commit a violent offence post-release. With respect to those who participated in the MIFVPP, there was evidence for significantly \((p < .01)\) reduced rates of violent recidivism as the intent-to-treat group was 57% less likely to commit a violent offence post-release (HR = 0.43 and 1.00 for intent-to-treat and comparison groups, respectively). While these two evaluations of the FVPP found some evidence for reduced recidivism rates as a result of treatment dosage, the results were not consistent across intensity levels. Notwithstanding these findings, it should be noted that while the ODARA and DVRAG can be used in assigning treatment intensity, given the static/historical nature of their items (and focus on the behaviour of the offender during the index assault), an offender’s participation in treatment would have little to no impact on the level of risk as assessed by either of the two measures. Therefore, it remains to be seen as to whether participation within the FVPP will have any detrimental impact on the recidivism base rate and predictive accuracy. Other factors such as ambiguity in victim identification (i.e., intimate partner
vs. non-intimate partner victim) have been found to underestimate base rates and detrimentally impact predictive validity (Hilton & Harris, 2009).

For the purpose of this study, in order to determine the minimum sample size required to ensure that a large AUC value (approximately .714; Rice & Harris, 2005) would be statistically significant at the $p < .05$ level, sample size calculations were conducted using statistical procedures described by Hanley and McNeil (1982; see Appendix B for SPSS syntax created by the author). However, rather than use a value of .714, the largest mean weighted AUC value from the meta-analysis previously described in Table 3 was chosen. Thus the weighted AUC value reported for the DVRAG (i.e., $AUC_w = .709$) was selected as a conservative estimate of a large AUC value. Of note is that the lower the AUC value, the larger the sample needed for statistical significance (Hanley & McNeil, 1982). Second, power was set to 80% ($Z_\beta = 0.84$), whereas significance was set to a one-sided 5% test of significance ($Z_a = 1.645$). Substituting these values into Formulas 2 and 3 from Hanley and McNeil (1982), the sample size required for an estimated AUC value of .709 to successfully reject the null hypothesis (i.e., $AUC = .50$) is $n = 90$. Of note is that the sample size formula (i.e., Formula 3) assumes a 50% base rate, therefore, to ensure that the estimated AUC value would retain significance the Wilcoxon standard error, $SE(W)$, was calculated for the following hypothetical base rates applied to a sample of $n = 90$: 10%, 20%, and 30%. Results of the analyses indicated that, while the standard error and, hence, statistical significance is influenced by base rate (i.e., as the base rate increases there is a reduction in the size of the standard error), the significance of the estimated AUC value would be retained at all hypothetical base rates: 10% ($AUC = .709$, $SE(W) = .101$, 95% CI [.511, .907], $p = .038$);
20% (AUC = .709, SE[\(W\)] = .074, 95% CI [0.564, 0.854], \(p = .005\)); and 30% (AUC = .709, SE[\(W\)] = .063, 95% CI [0.586, 0.832], \(p = .001\)). Therefore, a sample size of \(n = 90\) was determined to be the minimum required sample size for the current study.
Method

Participants

A total of 94 participants were drawn from a population of male federal offenders within the Ontario region. Within Canada the criteria for being incarcerated in a federal facility is receiving a custodial sentence of 2 years or more. For inclusion within the current study, the following three criteria were required: (1) the offender met the referral criteria for the FVPP, (2) there is documented proof (i.e., charges, convictions, police reports) that prior to their federal incarceration the offender committed IPV against a current or former female martial, cohabiting, or dating partner\(^4\), and (3) the offender has since completed his federal sentence (i.e., reached warrant expiry) or has been released from federal custody into the community on some form of structured release (i.e., unescorted temporary absence, day parole, full parole, or statutory release).

Measures and Recidivism Variables

The Domestic Violence Risk Appraisal Guide (DVRAG; Hilton et al., 2008; Hilton et al., 2010). The DVRAG is a 14-item actuarial risk assessment measure designed to predict domestic violence recidivism that incorporates all 13 items from the ODARA in addition to the PCL-R total score. Items are weighted based on their ability during the development phase to distinguish IPV recidivists from nonrecidivists. Scores can range from -10 to +46, with higher scores indicating elevated levels of risk, and can fall within one of 7 risk categories which provide an estimate for the probability of IPV

\(^4\) Although the ODARA and DVRAG were designed to assess risk for IPV among offenders who have assaulted a current or former marital or cohabiting partner, dating partners have since been included in the definition of the index offence victim (see Hilton, Harris, Popham et al., 2010, p. 817). Presently, there is no empirical or theoretical evidence to suggest that the application of the ODARA or DVRAG to dating violence cases would reduce its predictive validity (N. Z. Hilton, personal communication, November 25, 2011).
recidivism. High inter-rater reliability was found for DVRAG scores among 10 randomly
selected cases within the development sample (ICC = .90 [absolute agreement]), while
the calculated SEM based on data from the development sample was found to be 2.20
(Hilton et al., 2008). To date, no study has been conducted validating the DVRAG with
incarcerated offenders.

The Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004;
Hilton et al., 2010). The ODARA is a brief 13-item actuarial risk assessment measure
designed to assess risk for IPV recidivism. Items are scored in a dichotomous yes/no
format and can be completed by correctional staff in addition to frontline law personnel
following a review of the offender’s criminal history and information pertaining to the
index offence. Scores can range from 0 to 13, with higher scores indicating elevated
levels of risk, and can fall within one of 7 risk categories which provide an estimate for
the probability of IPV recidivism. High inter-rater reliability was found for ODARA
scores among 24 randomly selected cases within the development sample (ICC = .90
[absolute agreement]), while the calculated SEM based on data from the development
sample was found to be 0.94 (Hilton et al., 2004).

The Psychopathy Checklist-Revised (PCL-R; Hare, 2003). The PCL-R is a 20-
item rating scale designed to assess psychopathic personality traits. Scoring of the PCL-R
is based a semi-structured interview and thorough file review. Criteria for completing the
PCL-R are provided in detail within the user manual (Hare, 2003). Items are scored on a
3-point scale: 0 = does not apply, 1 = maybe/in some respect, and 2 = does apply.
Summing the 20-items produces a total score that can range from 0 to 40. Hare (2003)
has designated a score of 30 and above as the threshold for diagnosing psychopathy.
Factor analysis has identified a 2-factor, 4-facet model of psychopathy. According to the PCL-R manual\textsuperscript{5}, Factor 1 represents a "[s]elfish, callous, and remorseless use of others" (p. 79) encompassing Facets 1 (Interpersonal; Cronbach’s $\alpha = .71$) and 2 (Affective; Cronbach’s $\alpha = .71$), while Factor 2 represents a "[c]hronically unstable and antisocial lifestyle" (p. 79) encompassing Facets 3 (Lifestyle; Cronbach’s $\alpha = .67$) and 4 (Antisocial; Cronbach’s $\alpha = .64$). With respect to inter-rater reliability within research settings, Hare (2003) reports ICCs ranging from .67 to .86 for the PCL-R total and Facet scores among male offenders. Recently, Loomis and Ismail (2012) reported the ICCs for the PCL-R total and Facet scores assessed within an applied setting. They found ICCs ranging from .76 to .90 for the PCL-R total and Facet scores for a Canadian sample of 175 federally incarcerated male sexual offenders initially assessed during intake into federal custody (at Millhaven Assessment Unit [MAU]) and again following transfer to a federal psychiatric facility (i.e., Regional Treatment Centre of Ontario).

*The Spousal Assault Risk Assessment Guide 2nd edition (SARA; Kropp et al., 1999).* The SARA is a 20-item SPJ guideline comprised of static and dynamic risk factors related to spousal assault recidivism. There are four main components within the SARA represented by Criminal History (3 items related to historical risk factors), Psychosocial Adjustment (6 items related to social maladjustment), Spousal Assault History (7 items related to previous incidents of spousal assault), and Alleged/Most Recent Offense (3 items related to the current, most recent, or alleged index offense). In addition, there is an Other Considerations subscale within the SARA which consists of scores for risk factors which are not considered in the SARA but are deemed pertinent by the parole officer (e.g., pedophilia, sexual sadism). Items are scored on a 3-point scale: 0 = absent, 1 = sub-

\textsuperscript{5} Cronbach’s alpha coefficients as listed within the PCL-R manual for male offenders.
threshold, and 2 = present. Although the overall risk designation (low, moderate, or high) may be determined by the clinician, according to the manual, scores can be derived by summing the items for the general violence risk factors in Part 1 (Items 1 through 10; Cronbach’s $\alpha = .66$, ICC = .68), the spousal violence risk factors in Part 2 (Items 11 through 20; Cronbach’s $\alpha = .73$, ICC = .87), and all 20 items for the SARA total (Cronbach’s $\alpha = .78$, ICC = .84). Within the manual (Kropp et al., 1999), mean scores on the SARA were found to be significantly higher among incarcerated federal offenders ($n = 638$, $M = 16.39$, $SD = 6.86$) in comparison to offenders serving probation terms of 2 years or less ($n = 1,671$, $M = 12.98$, $SD = 6.46$); however, these differences were only attributed to significant differences in Part 1 of the SARA (i.e., the general violence risk factors).

The Statistical Information on Recidivism-Revised 1 (SIR-R1; Nafekh & Motiuk, 2002; Nuffield, 1982). The SIR-R1 is a 15-item actuarial risk measure currently used by the Correctional Service of Canada (CSC) for the purpose of institutional placement and release decision making. Each of the 15 items was associated with re-arrests in a sample of approximately 2500 offenders. Items found on the SIR-R1 are of a static nature and include type of current offence, age of admission, previous incarceration and history of violence. The majority of the items (12) relate to criminal history while the remaining items are related to marital status, number of dependents, and employment status at time of arrest. Scores range from -28 to 26, with negative scores indicating increased levels of risk. For the purpose of the present study, however, SIR-R1 scores were reversed such that higher scores were indicative of elevated levels of risk. Within
research contexts, the SIR-R1 has been found to have high inter-rater reliability ($r_{class} = .90$; Kroner, Mills, & Morgan, 2007).

**Recidivism and Revocation of Conditional Release.** Follow-up was conducted for all offenders following their release from federal custody. Correctional files (available on the Offender Management System [OMS]) and official police records (Finger Print Service [FPS] records of the Royal Canadian Mounted Police [RCMP]) were examined for the presence of violent recidivism. Violent recidivism was coded as a dichotomous variable (yes/no) on the basis of an offender receiving either a charge or conviction, or both. Included within the definition of violence were convictions or charges for robbery, assault, murder (includes manslaughter and attempted murder), contact sexual offences, kidnapping, arson with disregard for human life, assaulting a police officer, and uttering threats/criminal harassment. Violent offences committed within the context of an intimate relationship were coded as IPV recidivism in the form of both a dichotomous yes/no variable, two continuous variables previously coded by Hilton, Harris, Popham, and colleagues (2010): level of IPV severity (0 = no incidents, 1 = violation of conditional release, 2 = stalking or harassment, 3 = threat to physically harm or kill, and 4 = physical violence) and justice outcome (0 = no incidents, 1 = no police contact, 2 = police contact, 3 = charged, and 4 = convicted). Also recorded were offences detailed on OMS which do not appear on the CPIC but have sufficient information with trial and/or sentencing outcome. Charges were coded as recidivism only if the charge was dismissed, withdrawn, or received a stay of proceedings or peace bond. In the event that the offender was acquitted or found not guilty, the charges were not recorded in light of the offender being exonerated by the courts. Furthermore, as IPV can also occur while an offender is
incarcerated (via conjugal visits; Dutton & Hart, 1992), any offences committed while the offender was in custody, whether during federal or provincial incarceration, that garnered a criminal charge or conviction were included. Additionally, the offenders’ performance on structured release was coded based on whether they received a revocation of conditional release (dichotomous yes/no variable) as determined by the Parole Board of Canada (PBC).

**Time-at-risk.** Time-at-risk was the calculated number of days between the initial date of release from a federal institution and one of three subsequent dates: date of offence/conviction/charge, date of death, or date of follow-up (i.e., between May 15, 2012 and June 22, 2012). This variable was only calculated for the violent and IPV dichotomous recidivism variables. Furthermore, in the event that an offender committed a violent or IPV offence prior to release (i.e., while incarcerated), his time at risk was the calculated number of days between the date of the family violence assessment and the date of the offence.

**Treatment dosage.** Treatment dosage was estimated using the total number of group sessions attended. However, as a result of missing or incomplete data, the calculated number of days between the start date and either the program completion date or the date of program suspension served as a proxy measure of the number of group sessions attended. Among the offenders in which the number of groups sessions attended was reported \((n = 37)\), the calculated number of days in program were found to be highly correlated with the recorded number of group sessions attended \((r = .94, p < .001)\).

**Procedure**
Prior to data collection, research approval was granted by the Psychology Research Ethics Board of Carleton University (Study No. 11-139; see Appendix C) and the Research Branch of the Correctional Service of Canada (see Appendix D). All of the offenders within the sample provided informed consent for their program information to be gathered and used for research purposes, prior to their participation in the FVPP.

**Data collection and scoring of the ODARA and DVRAG.** As stated earlier, the current study consists of a hybrid prospective-retrospective design. Data regarding the SARA and SIR-R1 scale were previously collected and used for research purposes (e.g., Connors et al., 2012; Connors et al., in press; Gray, Mills, & Connors, 2011). The SIR-R1 was completed by trained parole officers at the beginning of the offenders’ sentence as part of the offender intake assessment. Prior research using the SIR-R1 scale scored by parole officers during the intake assessment have found it to be moderately associated with recidivism \( r = .27 \); Kroner, Gray, & Goodin, 2011). Similarly, the SARA was completed by a trained parole officer as part of the family violence assessment if during the offender intake process there was sufficient evidence, based on official documentation (e.g., criminal record), that an offender was a known perpetrator, victim, or suspected perpetrator of IPV. Of note is that the presence of IPV was not solely determined based on charges and convictions recorded on criminal records, but may have also incorporated police reports and self-report from the offender or a family member. Despite the presence of dynamic risk factors on the SARA, scoring of the SARA only occurred prior to treatment as its scores and summary risk ratings were used to inform treatment intensity as opposed to assessing offender change resulting from treatment or life circumstances.
To facilitate the reliable scoring of the ODARA and DVRAG, prior to data collection, scoring certification for the current author was achieved for the ODARA by scoring a total of \( n = 10 \) cases provided within the *ODARA 101: The Electronic Training Program* (Waypoint Centre for Mental Health Care, 2011) available in interactive DVD format (Certificate No. 12-083; see Appendix F). Level of inter-rater agreement between the ODARA risk categories for the 10 cases as independently scored by the lead developer (N. Zoe Hilton) and the current author was considered high (single measures, random effects, absolute agreement; ICC = .978, \( p < .001 \)).

Within the current study, scoring of the ODARA and DVRAG items was based on a thorough review of correctional file information pertaining to psychological risk assessments, offender intake assessments, criminal profiles, and FVPP reports, in addition to various police and court documents located within the Police and Court Information Management Module (PCIMM) available through OMS. Furthermore, hard copy PCL-R item, facet, and total scores were collected from the psychology files located at the Offender Records Depot within Millhaven Institution. In the event that there were two independently rated PCL-R assessments available on file, information from both assessments was recorded. This allowed for an evaluation of the inter-rater reliability of the PCL-R scores between independent raters, so long as the offender was not released nor incurred any new charges or convictions between the two assessments. Moreover, in some instances where hard copy PCL-R information was unavailable, PCL-R total, factor, and facet scores were derived from percentile rankings reported within the psychological risk assessments obtained from the OMS. Lastly, level of victim injury at
the IPV index offence was coded based on the item from the VRAG (-2 = death, 0 = hospitalized, +1 = treated and released, and +2 = none or slight [includes no victim]).

Data Analysis

Frequency distributions and descriptive statistics (i.e., means and standard deviations) were computed for the ODARA and DVRAG to determine the level of item endorsement and to compare the norms of the current federal offender sample to those previously reported by Hilton and colleagues (Hilton et al., 2008; Hilton, Harris, Popham et al., 2010) through determining the magnitude of the difference (i.e., Cohen's \( d \)). To examine inter-rater reliability of the ODARA and DVRAG, Pearson's \( r \) and ICC (single measures, two-way mixed, absolute agreement; and single measures, two-way mixed, consistency) were calculated for 10 randomly selected cases, whereas structural reliability was examined by calculating the corrected item-total correlations and Cronbach's \( \alpha \).

To examine the concurrent and discriminant validity of the ODARA and DVRAG, correlational analyses were conducted to ascertain the level of associations between the two IPV risk assessment measures and that of the SIR-R1, SARA, and PCL-R. Furthermore, predictive validity of the five risk assessment measures was determined by examining their association with the dichotomous recidivism variable via calculated point-biserial correlations, AUCs, and Cohen's \( d \) values; whereas only correlational analyses were conducted between the total scores and the level of severity for IPV recidivism and justice outcome. Furthermore, to determine whether treatment dosage had a detrimental impact on the predictive validity of the measures, partial point-biserial correlations (partialling out treatment dosage) were calculated between the risk
assessment measures and the dichotomous IPV recidivism and continuous IPV severity variables.

Finally, for comparative purposes and to control for variations in time-at-risk (Hanson, 2009), Cox regression survival analysis (Cox, 1972) was used to examine the unique contribution that the PCL-R, SARA, and SIR-R1 make to the prediction of any violent recidivism, and more specifically IPV recidivism relative to the ODARA. In doing so, the ODARA was first entered into a hierarchical Cox regression at Block 1, whereas the PCL-R, SARA, and SIR-R1 scores were entered at Block 2, with time-at-risk entered as the time variable and dichotomous violent and IPV recidivism as the status. Of note is that the DVRAG was excluded from this analysis due to the increased risk of multicolinearity with the ODARA and PCL-R.
Results

Sample Characteristics

Of the sample of 586 offenders previously identified (Connors et al., 2012, in press), a total of 224 were identified as having served their federal sentence within Ontario. Offenders were then screened based on whether they met the criteria for a psychological risk assessment to be conducted during one of the following three stages of their federal sentence: at intake, prior to transfer to lower security, or prior to community release. As outlined by CSC within the Commissioner's Directives (CD 712-1), a psychological risk assessment is mandatory if an offender meets any of the following criteria: (a) exhibits persistent or gratuitous violence, (b) is referred for detention (i.e., detained until warrant expiry), (c) requires a conditional release review as a result of serving an indeterminate sentence (includes those serving life sentences and those declared as a dangerous offender [DO]), or (d), is a sex offender previously identified within the Specialized Sex Offender Assessment as being either a high risk to reoffend, or is considered a moderate risk offender who remains untreated or has dropped out of a core correctional treatment program.

Among the 224 offenders, 146 (65.2%) had psychological risk assessments on file in which the PCL-R was administered. Total SARA and SIR-R1 scores were available for 121 of the 146 offenders and among the 121, 94 met the three criteria for inclusion and had sufficient information to score the ODARA and DVRAG. Table 5 displays the sample characteristics for the current study. Of note was that, while all 94 offenders had been scored on the PCL-R, only 91 PCL-R scores could be derived through reported percentile ranks (n = 58, 63.7%) or hard copy file information available on their psychology files (n = 33, 36.3%). Additionally, for those serving determinate sentences,
Table 5

Sample Characteristics

<table>
<thead>
<tr>
<th>Sample characteristic</th>
<th>M (SD)</th>
<th>Range</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>35.58 (9.13)</td>
<td>20 – 57</td>
<td>94 (100.0)</td>
</tr>
<tr>
<td>Assessed risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCL-R score</td>
<td>21.23 (5.85)</td>
<td>4 – 35</td>
<td>91 (96.8)</td>
</tr>
<tr>
<td>SARA score</td>
<td>18.15 (5.91)</td>
<td>5 – 32</td>
<td>94 (100.0)</td>
</tr>
<tr>
<td>SIR-R1 score</td>
<td>5.33 (7.51)</td>
<td>-23 – 18</td>
<td>94 (100.0)</td>
</tr>
<tr>
<td>Victim injury</td>
<td>1.43 (0.89)</td>
<td>-2 – 2</td>
<td>94 (100.0)</td>
</tr>
<tr>
<td>Death</td>
<td>3 (3.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalized</td>
<td>7 (7.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated and released</td>
<td>28 (29.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None or slight</td>
<td>56 (59.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community supervision</td>
<td>9 (9.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceased (within the community)</td>
<td>1 (1.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Re-incarcerated</td>
<td>12 (12.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence complete</td>
<td>72 (76.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVPP intensity level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>33 (35.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>61 (64.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPV related federal index offence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>56 (59.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>38 (40.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racial composition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aboriginal</td>
<td>1 (1.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African Canadian</td>
<td>6 (6.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>79 (84.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>8 (8.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship with current/former partner</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common-law/cohabiting</td>
<td>53 (56.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dating</td>
<td>26 (27.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>15 (16.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence type/length (months)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determinate</td>
<td>40.68 (24.49)</td>
<td>10 – 120</td>
<td>88 (93.6)</td>
</tr>
<tr>
<td>Determinate with LTSO</td>
<td>51.25 (34.48)</td>
<td>0 – 73</td>
<td>4 (4.3)</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>2 (2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of IPV offender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stalker</td>
<td>2 (2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent</td>
<td>92 (97.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. M = mean; SD = standard deviation; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1; FVPP = Family Violence Prevention Program; LTSO = Long-term supervision order.
the calculated sentence length (in months) ranged from 0 to 120. This was due to a number of cases ($n = 22$) serving sentences of 23 months or less which falls below the 2 year cut-off for federal incarceration. Closer inspection revealed that sentence length was reduced for a number of the offenders through various appeal processes thus shortening the time between the commencement of their sentence and warrant expiry. One case was identified as receiving a sentence of time served in pre-trial custody resulting in his calculated sentence length equaling 0 months. However, this offender came under federal jurisdiction due to receiving a long-term supervision order (LTSO) greater than 10 years.

With respect to FVPP treatment, approximately $35.1\% (n = 33)$ were referred to the high intensity program (HIFVPP), whereas the remaining $64.9\% (n = 61)$ were referred to the moderate (MIFVPP; low risk offenders are referred to the Family Violence Awareness Program). Of the 94 referred to treatment, 89 entered the treatment program. Treatment dosage ranged from 2 to 183 days ($M = 83.10, SD = 42.63$). Sixty of the offenders successfully completed treatment (HIVPP = 19, MIFVPP = 41), 4 attended all sessions (HIVPP = 3, MIFVPP = 1), 10 did not complete (HIVPP = 1, MIFVPP = 9), and 15 were suspended (HIVPP = 6, MIFVPP = 9). The remaining 5 who were initially referred ultimately did not enter into the FVPP due to being referred to other programs (e.g., anger and emotions, sex offender, or violence prevention programming), unavailability of the program due to wait-listing and time constraints, or being evaluated by the program coordinators as not meeting the program criteria (i.e., denial of prior IPV offences). Those whose index federal offence was not IPV related were at increased likelihood of being referred to the MIFVPP ($\chi^2 [1] = 11.38, p = .001$). Comparisons between the two groups on the SARA subscales and SIR-R1 are described in Appendix
F. No significant differences were noted between offenders referred to the HIVPP in comparison to those referred to the MIFVPP on the PCL-R total score ($t_{[89]} = 0.69, p > .05$). Similarly, no meaningfully significant differences were noted between the two groups on $F_1 (t_{[86]} = 0.03, p > .05)$ or $F_2 (t_{[87]} = 0.89, p > .05)$ of the PCL-R.

**Frequency Distributions and Reliability Analyses**

Among the 94 cases with sufficient information to score the ODARA, 36 (38.3%) had one missing item, whereas two single cases had 2 and 3 missing items, respectively. Table 6 displays the frequency distributions for the 13 dichotomous ODARA items, in addition to the 6 continuous (or count) items from the DVRAG. Victim concern was the most problematic as 39 of the cases could not be scored. Moreover, prorating of the ODARA scores was deemed inappropriate as the majority of the cases with missing items ($n = 27, 71\%$) either met or exceeded the 7+ category prior to prorating. Additionally, only a single case with more than one missing item fell below the 7+ category (i.e., the score was equal to 6). Therefore, all analyses with the ODARA used raw scores with missing information being scored as a 0. Despite the presence of missing information, 73 (77.7%) of the offenders within the current sample had ODARA scores which fell within the highest risk category (i.e., 7-13). The overall mean score on the ODARA among the 94 offenders was 8.06 ($SD = 2.05$, range 2 to 12) which was significantly higher, $t(93) = 10.65, p < .001$, Cohen's $d = 1.09$, in comparison to the mean of the incarcerated sample ($M = 5.81, SD = 2.06, N = 150$) previously reported by Hilton, Harris, Popham et al. (2010). Whereas prorating was not feasible with the ODARA, all completed DVRAGs with missing information were prorated as outlined by Hilton et al. (2010). The overall mean score on the DVRAG was 25.59 ($SD = 9.13$, range -7 to +40) which was
significantly higher, \( t(93) = 24.10, p < .001 \), Cohen's \( d = 2.86 \), in comparison to the mean of the development sample (\( M = 2.88, SD = 7.76, N = 649; \) Hilton et al., 2010). The mean DVRAG risk category was 6.53 (\( SD = 0.89 \), range 2 to 7) with 28.7% (\( n = 27 \)) and 66.0% (\( n = 62 \)) falling into the second highest (i.e., risk category 6) and highest (i.e., risk category 7) risk categories, respectively. No significant differences were noted between the MIFVPP or HIVPP on either the ODARA (\( t[92] = 0.50, p > .05 \)) or DVRAG (\( t[92] = 0.05, p > .05 \)).

Results of the structural reliability analysis reported in Table 6 revealed poor internal consistency among the ODARA and DVRAG items (Cronbach's \( \alpha = .53 \) [\( n = 94 \)] and \( .48 \) [\( n = 89 \)], respectively). Although corrected item-total correlations were relatively stable across both measures with the average corrected item-total correlation for the ODARA and DVRAG equaling .21 and .18, respectively, visual inspection of the corrected item-total correlations indicated that the items relating to confinement of partner at index assault and assault on victim when pregnant were not necessarily consistent with the other items. A caveat, however, is that the level of internal consistency exhibited by the two measures is undoubtedly biased as the cases with missing victim concern scores were scored as a 0 on the item for both the ODARA and DVRAG to allow for the corrected item-total correlations and Cronbach's \( \alpha \) calculations to be made with the largest sample possible. While these methods of assessing reliability reflect the Classical Test Theory (Kropp et al., 1999), it should be noted that they are rarely applied to risk assessment measures and are even considered "irrelevant because the instrument was not intended to measure an underlying [psychological] construct" (Douglas, Skeem, & Nicholson, 2011, p. 334). Therefore, these results are only tentative.
Table 6
Item Endorsement and Corrected Item-Total Correlations ($r_{it}$) for the ODARA and DVRAG

<table>
<thead>
<tr>
<th>Item Endorsement (%)</th>
<th>Item</th>
<th>ODARA</th>
<th>DVRAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing 0</td>
<td>1</td>
<td>$r_{it}$</td>
<td>$r_{it}$</td>
</tr>
<tr>
<td>Prior domestic incident</td>
<td>No. of assaults</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = 18.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = 19.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2 = 62.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior nondomestic incident</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0 = 18.1</td>
<td>81.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 = 91.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2 = 81.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior custodial sentence ≥ 30 days</td>
<td>0.0 = 8.5</td>
<td>91.5</td>
<td></td>
</tr>
<tr>
<td>Failure on prior conditional release</td>
<td>0.0 = 7.4</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>Threat to harm or kill at the index assault</td>
<td>1.0 = 1.1</td>
<td>55.3</td>
<td>43.6</td>
</tr>
<tr>
<td>Confinement of partner at index assault</td>
<td>1.1 = 1.1</td>
<td>71.3</td>
<td>27.7</td>
</tr>
<tr>
<td>Victim concern</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.4 = 40.4</td>
<td>47.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than one child</td>
<td>No. of children</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = 62.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = 20.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2 = 17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victim’s biological children from a previous partner</td>
<td>0.0 = 0.0</td>
<td>37.2</td>
<td></td>
</tr>
<tr>
<td>No. of children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = 62.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = 20.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2 = 17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violence against others</td>
<td>0.0 = 18.1</td>
<td>81.9</td>
<td></td>
</tr>
<tr>
<td>Substance abuse</td>
<td>0.0 = 13.8</td>
<td>86.2</td>
<td></td>
</tr>
<tr>
<td>No. of factors present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = 18.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = 81.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assault on victim when pregnant</td>
<td>0.0 = 0.0</td>
<td>92.6</td>
<td>7.4</td>
</tr>
<tr>
<td>Barriers to victim support</td>
<td>1.1 = 1.1</td>
<td>28.7</td>
<td>70.2</td>
</tr>
<tr>
<td>No. of factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = 28.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = 55.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2 = 14.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychopathy Checklist-Revised score</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 9 = 1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 – 16 = 20.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 17 = 75.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. DVRAG = Domestic Assault Risk Assessment Guide; ODARA = Ontario Domestic Assault Risk Assessment.
at best and may not necessarily reflect the true structural reliability of the two measures.

Inter-rater reliability analyses (see Table 7) between the current author and an independent rater with multiple publications and years of experience in assessing violence risk (J. Mills) were conducted on ODARA total scores and scores on the DVRAG continuous (or count) items for a random sample of 10 cases. To ensure unbiased scoring, the independent rater was blind to both the offenders’ ODARA and DVRAG scores (as scored by the current author) and recidivism data. Calculated Pearson’s $r$ and ICCs revealed significantly high levels (within the “excellent” range, ICC > .75; Fleiss, 1981) of inter-rater reliability for both the ODARA and DVRAG continuous items with respect to agreement between scores (i.e., absolute agreement) and ranking of cases (i.e., consistency). Similar levels of inter-rater reliability have been exhibited previously within studies utilizing the ODARA and DVRAG (Hilton et al., 2004, 2008; Rettenberger & Eher, 2012); however, unlike prior studies which have evaluated inter-rater reliability based on risk categories, this study evaluated agreement between total scores as there was no variability among risk categories. With respect to the remaining risk assessment measures, independent rater scores were available from file information for both the PCL-R and the SIR-R1 for 16 and 9 cases, respectively, among the larger sample ($n = 146$) that had psychological risk assessments on file in which the PCL-R was administered. Both measures exhibited high levels of inter-rater reliability with ICCs ranging from .909 to .860 for total scores. While the ICC for $F_1$ of the PCL-R was just below .70, it was within the “good” range (i.e., ICC of .60 to .74; Fleiss, 1981).
Table 7

Inter-Rater Reliability Analyses

<table>
<thead>
<tr>
<th>Measure</th>
<th>k</th>
<th>r</th>
<th>Absolute Agreement</th>
<th>Consistency</th>
<th>Score differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVRAG</td>
<td>10</td>
<td>.82*</td>
<td>.809*</td>
<td>.819*</td>
<td>0 – 5</td>
</tr>
<tr>
<td>ODARA</td>
<td>10</td>
<td>.95**</td>
<td>.901**</td>
<td>.927**</td>
<td>0 – 2</td>
</tr>
<tr>
<td>PCL-R</td>
<td>16</td>
<td>.92**</td>
<td>.909**</td>
<td>.907**</td>
<td>0 – 8</td>
</tr>
<tr>
<td>Factor 1</td>
<td>13</td>
<td>.70*</td>
<td>.693*</td>
<td>.677*</td>
<td>0 – 5</td>
</tr>
<tr>
<td>Factor 2</td>
<td>13</td>
<td>.98**</td>
<td>.965**</td>
<td>.976**</td>
<td>0 – 3</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>9</td>
<td>.86*</td>
<td>.860*</td>
<td>.851*</td>
<td>0 – 7</td>
</tr>
</tbody>
</table>

Note. k = number of ratings; r = correlation coefficient between independent rater scores; ICC = interclass correlation coefficient (single measures, two-way mixed); M = mean; SD = standard deviation; DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SIR-R1 = Statistical Information on Recidivism-Revised 1.

Concurrent and Discriminant Validity

Correlational analyses between the various risk assessment measures and the ODARA and DVRAG scores supported the concurrent and discriminant validity of the two measures (see Table 8). Similar to prior research, the correlation between the ODARA and DVRAG was large and significant (r = .82, p < .001). Despite significant associations previously found between the PCL-R and the ODARA, within the current study the correlation between the two was small and non-significant (r = .12, p > .05). Similarly, the ODARA appeared to be unrelated to F1, yet significantly related to F2 of the PCL-R. Interestingly, the ODARA was moderately and significantly associated with both the SARA and SIR-R1 total scores (r = .35 and .34, respectively). The strong association with the SARA appeared to be related to the Criminal History, Psychosocial Adjustment, and Spousal Assault History subscales of the SARA. Not surprisingly the
DVRAG was significantly and moderately associated with the PCL-R total and $F_2$ scores. While the $F_1$ score was only modestly related to the DVRAG, it did appear to be approaching statistical significance ($r = .20, p = .056$). Moreover, the association between the DVRAG and both the SARA total and Spousal Assault History scores was much smaller in comparison to the ODARA. This is likely due to variance relating to general criminal risk being introduced into the DVRAG score by way of its incorporation of the PCL-R. Indeed, this observation is supported by the substantial increase in association between the DVRAG and the three general criminal risk measures (i.e., $F_2$ of the PCL-R, Criminal History subscale of the SARA, and the SIR-R1). While both the ODARA and DVRAG are accounting for IPV risk, the latter appears to be accounting for increased general criminal risk variance and less IPV risk variance in comparison to the former.

Both the ODARA and DVRAG were moderately significantly related to the level of victim injury inflicted during the index offence such that lower risk offenders caused increased victim injury. Similar results were found between victim injury and the SIR-R1 ($r = .37, p < .001$) and Criminal History subscale of the SARA ($r = .37, p < .001$). However, the reverse was found between victim injury and the Alleged/Most Recent Offense subscale of the SARA ($r = -.26, p = .013$) and, to a lesser extent, the Other Considerations subscale of the SARA ($r = -.18, p = .085$). Correlations between the PCL-R, SARA, and SIR-R1 are reported in Appendix G. Overall, while the results of the correlational analyses indicated that the ODARA and DVRAG are moderately related to the general and IPV risk measures, they are clearly accounting for variance not captured by these other measures.
Table 8

Correlations between the ODARA, DVRAG and the PCL-R, SARA, SIR-R1, and Victim Injury

<table>
<thead>
<tr>
<th>Measure</th>
<th>ODARA score</th>
<th>DVRAG score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychopathy Checklist-Revised</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 1</td>
<td>.12</td>
<td>.39***</td>
</tr>
<tr>
<td>Factor 2</td>
<td>.02</td>
<td>.20†</td>
</tr>
<tr>
<td>Spousal Assault Risk Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criminal history</td>
<td>.24*</td>
<td>.43***</td>
</tr>
<tr>
<td>Psychosocial adjustment</td>
<td>.25*</td>
<td>.38***</td>
</tr>
<tr>
<td>Spousal assault history</td>
<td>.28**</td>
<td>.28**</td>
</tr>
<tr>
<td>Alleged/most recent offense</td>
<td>.23*</td>
<td>.16</td>
</tr>
<tr>
<td>Other considerations</td>
<td>.09</td>
<td>-.11</td>
</tr>
<tr>
<td>Statistical Information on Recidivism-Revised</td>
<td>.05</td>
<td>-.01</td>
</tr>
<tr>
<td>Victim injury (index offense only)</td>
<td>.34**</td>
<td>.52***</td>
</tr>
<tr>
<td></td>
<td>.31**</td>
<td>.35**</td>
</tr>
</tbody>
</table>

Note. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment.
†p < .10, *p < .05, **p < .01, ***p < .001.

Recidivism Base Rates and Predictive Validity Analyses

The calculated follow-up time between the family violence assessment and end of the follow-up period for the current sample ranged from 41 to 120 months (M = 65.04 months, SD = 14.06). Of the 94 offenders, 84 had surpassed their warrant expiry by approximately 1 to 66 months (M = 32.81 months, SD = 14.35). Average time since release was 43.62 months (SD = 15.03, range = 1 to 86 months). Among the offenders released into the community under federal supervision (n = 88), over half (n = 49) had their release revoked by the PBC. The reduction in sample size was due to a small number of offenders being detained until warrant expiry or receiving immediate release with time served due to various appeals. Predictive analyses concerning revocation of conditional release revealed non-significant point-biserial correlations and AUC values.
among the risk assessment measure total and subscale scores (see Appendix H). Hence, no further statistical analyses were conducted in this regard.

Recidivism base rates for the sample were as follows: 43.6% for any recidivism ($n = 41$), 23.4% for violent recidivism ($n = 22$), and 12.8% for IPV recidivism ($n = 12$). Among the IPV recidivists, 5 were considered ambiguous as the relationship between the offender and victim could not be ascertained. Regardless, these 5 cases were included as IPV recidivists given their pattern of offending behaviour and based on the nature and location of the charge or conviction. Among the recidivists, 2 were convicted for stalking or harassment, 1 was charged and 1 was convicted of threat to physically harm or kill, and 2 were charged while 6 were convicted of physical violence. Of the 5 cases in which the IPV victim could be identified, 3 offended against the same victim as the index assault and 2 offended against a new intimate partner. Including IPV offences, violent offences (based on charges and convictions) committed by the 22 offenders during the follow-up period were intimidation ($n = 1$), criminal harassment ($n = 1$), uttering threats ($n = 3$), robbery ($n = 5$), assaulting a peace officer ($n = 1$), assault ($n = 6$), assault with a weapon ($n = 3$), and sexual assault ($n = 2$). Table 9 and 10 display the results of the predictive validity analyses (i.e., $r_{pb}$, Cohen's $d$, and AUC analyses) for violent and any recidivism, and IPV recidivism outcomes, respectively.
### Table 9
**Accuracy of Five Risk Assessment Measures in Predicting Violent Recidivism and Any Recidivism (N = 94)**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Violent recidivism (23.4%)</th>
<th>Any recidivism (43.6%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_{pb}$</td>
<td>$d$</td>
</tr>
<tr>
<td>DVRAG</td>
<td>.29**</td>
<td>.72</td>
</tr>
<tr>
<td>ODARA</td>
<td>.19†</td>
<td>.45</td>
</tr>
<tr>
<td>PCL-R&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.20†</td>
<td>.48</td>
</tr>
<tr>
<td>Factor 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.29**</td>
<td>.68</td>
</tr>
<tr>
<td>Factor 2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>SARA</td>
<td>.12</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychosocial adjustment</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Spousal assault history</td>
<td>.15</td>
<td>.35</td>
</tr>
<tr>
<td>Alleged/most recent offense</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>Other considerations</td>
<td>-.04</td>
<td>-.09</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>.15</td>
<td>.35</td>
</tr>
</tbody>
</table>

*Note.* DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1. AUC = Area under the curve of the receiver operating characteristic (ROC); 95% CI<sub>BE</sub> = Binomial exact 95% confidence intervals of the AUC based on standard error derived from the DeLong et al. (1988) method calculated using MedCalc; 95% CI<sub>NP</sub> = Nonparametric 95% confidence interval of the AUC based on standard error derived from the Hanley and McNeil (1982) method calculated using SPSS. Base rates for violent and any recidivism are in parentheses. Significance of the AUC values for violent recidivism was determined through a z-test (i.e., $z = [\text{AUC} - .5]/\text{standard error}$) and is therefore not based on the 95% CI<sub>BE</sub>.  
<sup>a</sup>N = 91. Base rate for violent and any recidivism = 24.2% and 45.1%, respectively.  
<sup>b</sup>N = 88. Base rate for violent and any recidivism = 25.0% and 44.3%, respectively.  
<sup>c</sup>N = 89. Base rate for violent and any recidivism = 24.7% and 44.9%, respectively.  
†$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.  

---
Table 10
Accuracy of Five Risk Assessment Measures in Predicting IPV Recidivism Outcomes (N = 94)

<table>
<thead>
<tr>
<th>Measure</th>
<th>IPV recidivism (12.8%)</th>
<th>IPV severity</th>
<th>Justice outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_{pb}$</td>
<td>$d$</td>
<td>AUC</td>
</tr>
<tr>
<td>DVRAG</td>
<td>.22*</td>
<td>.67</td>
<td>.713***</td>
</tr>
<tr>
<td>ODARA</td>
<td>.25*</td>
<td>.77</td>
<td>.712**</td>
</tr>
<tr>
<td>PCL-R$^a$</td>
<td>.12</td>
<td>.35</td>
<td>.607</td>
</tr>
<tr>
<td>Factor 1$^b$</td>
<td>.20†</td>
<td>.60</td>
<td>.685**</td>
</tr>
<tr>
<td>Factor 2$^c$</td>
<td>-.02</td>
<td>-.06</td>
<td>.529</td>
</tr>
<tr>
<td>SARA</td>
<td>.16</td>
<td>.48</td>
<td>.604</td>
</tr>
<tr>
<td>Criminal history</td>
<td>.02</td>
<td>.07</td>
<td>.508</td>
</tr>
<tr>
<td>Psychosocial adjustment</td>
<td>.06</td>
<td>.18</td>
<td>.583</td>
</tr>
<tr>
<td>Spousal assault history</td>
<td>.18†</td>
<td>.53</td>
<td>.635</td>
</tr>
<tr>
<td>Alleged/most recent offense</td>
<td>.08</td>
<td>.25</td>
<td>.575</td>
</tr>
<tr>
<td>Other considerations</td>
<td>.02</td>
<td>.05</td>
<td>.532</td>
</tr>
<tr>
<td>SIR-R$^l$</td>
<td>.09</td>
<td>.27</td>
<td>.590</td>
</tr>
</tbody>
</table>

Note. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R = Statistical Information on Recidivism-Revised

1. AUC = Area under the curve of the receiver operating characteristic (ROC); 95% CI BE = Binomial exact 95% confidence intervals of the AUC based on standard error derived from the DeLong et al. (1988) method. Base rate for IPV recidivism is in parentheses. Significance of the AUC values for IPV recidivism was determined through a z-test (i.e., $z = \frac{[AUC - .5]}{\text{standard error}}$) and is therefore not based on the 95% CI BE.

$^a$N = 91. Base rate for IPV recidivism = 13.2%.

$^b$N = 88. Base rate for IPV recidivism = 13.6%.

$^c$N = 89. Base rate for IPV recidivism = 13.5%.

$^†p < .10, * p < .05, ** p < .01, *** p < .001.$
Given that the main focus of the current study was on violent and IPV recidivism, these are the only results discussed, and, hence, no comparative analyses were conducted with the any recidivism outcome variable. It is noted, however, that the SIR-R1 displayed the largest AUC value (.628, \( p < .05 \)), followed by the DVRAG (.623, \( p < .05 \)), and lastly \( F_1 \) of the PCL-R (.606, \( p = .089 \)). Moreover, as displayed within Table 9 and Table 10, the DVRAG produced the largest AUC values in predicting violent and IPV recidivism (AUC = .708, \( p < .001 \), and .713, \( p < .001 \), respectively). While the DVRAG produced the largest AUC values, the ODARA produced the largest Cohen's \( d \) value for IPV recidivism. This is likely due to the differences in calculation between the two effect sizes as the AUC value is a rank order statistic whereas Cohen's \( d \) represents the magnitude of the difference in mean scores between the recidivists and non-recidivists. \( F_1 \) of the PCL-R was a significant predictor of both violent and IPV recidivism (AUC = .686, \( p < .01 \), and .685, \( p < .01 \), respectively), whereas the Spousal Assault History subscale of the SARA displayed moderate predictive accuracy with respect to IPV recidivism (AUC = .635, \( d = .53 \)) yet did not achieve statistical significance (i.e., \( z = 1.35 \)). Whereas the DVRAG and ODARA displayed significantly modest associations with IPV severity and justice outcome, among the remaining measures, only \( F_1 \) of the PCL-R and the Spousal Assault History subscale of the SARA were within trend level significance with respect to their associations with the justice outcome variable.

Due to the relatively low recidivism base rates for violent and IPV recidivism, the exact binomial 95% confidence intervals (based on the standard error derived from the DeLong, DeLong, and Clarke-Pearson [1988] method) were calculated for each measure using MedCalc statistical software (these confidence intervals were not calculated for any
recidivism as the base rate was nearing 50%). Furthermore, the nonparametric 95% confidence intervals calculated by SPSS are based on the standard error derived from Hanley and McNeil's (1982) method. This is somewhat problematic as the Hanley and McNeil (1982) standard error is considered a conservative approximation which results in confidence intervals that are considered "excessively wide" on account of the inflated standard error (Brown & Davis, 2006, p. 30). Therefore, all post hoc comparative analyses between the AUC values derived by the risk assessment measures were conducted using the methods outlined by DeLong et al. (1988). Briefly, this method examines the difference between two AUC values derived from a single sample while taking into account the covariance between the two AUC values (DeLong et al., 1988; Hanley & Hajian-Tilaki, 1997). However, to conserve power for the comparative analyses, two sets of contrasts were conducted post hoc using only the total scores of the five measures (i.e., 10 comparisons per contrasting set). The first contrast consisted of the total scores for the five measures with respect to violent recidivism of which significant differences were found between the DVRAG and both the ODARA and SARA total scores (\(z = 1.98, p = .048\), and \(z = 2.08, p = .038\), respectively; Bonferroni adjustment for \(p \leq .05 = .005\)). The second contrast consisted of the total scores for the five measures with respect to IPV recidivism of which there were no significant differences noted\(^6\).

Next, to examine the impact that treatment dosage had on the predictive validity of the risk assessment measures, partial point-biserial and partial correlations (controlling for treatment dosage) were calculated for the 89 offenders who entered either the high or moderate intensity FVPP (see Table 11). While some minor reductions were noted in the associations for the DVRAG and ODARA with IPV recidivism and severity, they did not

\(^6\)Replacing the PCL-R total score with \(F_1\) within the first and second contrasts did not alter the results.
appear to significantly impact the predictive validity of the two measures. Indeed, this was noted for all of the measures and various subscales with the exception of the SARA total score, Spousal Assault History subscale, and Alleged/Most Recent Offence subscale in which accounting for treatment dosage increased the association with IPV recidivism and severity, particularly for the Spousal Assault History subscale. Regardless, these results did reveal that an offender’s participation in treatment has little to no impact on the level of risk as assessed by either the DVRAG or ODARA. Contrasting this, however,

Table 11
Partial Correlations Controlling for Treatment Dosage

<table>
<thead>
<tr>
<th>Measure</th>
<th>IPV recidivism</th>
<th>IPV severity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partial</td>
<td>Actual</td>
</tr>
<tr>
<td>DVRAG</td>
<td>89</td>
<td>.21*</td>
</tr>
<tr>
<td>ODARA</td>
<td>89</td>
<td>.25*</td>
</tr>
<tr>
<td>PCL-R</td>
<td>86</td>
<td>.11</td>
</tr>
<tr>
<td>Factor 1</td>
<td>83</td>
<td>.20†</td>
</tr>
<tr>
<td>Factor 2</td>
<td>84</td>
<td>-.01</td>
</tr>
<tr>
<td>SARA</td>
<td>89</td>
<td>.23*</td>
</tr>
<tr>
<td>Criminal history</td>
<td>89</td>
<td>.02</td>
</tr>
<tr>
<td>Psychosocial adjustment</td>
<td>89</td>
<td>.10</td>
</tr>
<tr>
<td>Spousal assault history</td>
<td>89</td>
<td>.24*</td>
</tr>
<tr>
<td>Alleged/most recent offense</td>
<td>89</td>
<td>.15</td>
</tr>
<tr>
<td>Other considerations</td>
<td>89</td>
<td>.04</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>89</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1. †p < .10, *p < .05.

is the result with respect to the Spousal Assault History subscale of the SARA. While the majority of the items within the subscale are static in nature, there are items that are
possibly being impacted by treatment. These items relate to the degree to which the
offender minimizes or denies his prior IPV offending behaviour and his level of
endorsement with respect to attitudes that support or condone IPV. Of note is that
attitudes supportive of crime are considered to be one of the most important criminogenic
needs identified within the PCC (Andrews & Bonta, 2010).

**Cox Regression Survival Analysis**

To account for variations in follow-up time and to assess the incremental
correlation of the various risk assessment measures in predicting recidivism, a series of
hierarchical Cox regression survival analyses were conducted (see Table 11). To
determine whether the PCL-R, SARA, and SIR-R1 added incrementally to the prediction
of violent recidivism relative to the ODARA, the ODARA total score was entered at
Block 1, whereas the PCL-R, SARA, and SIR-R1 total scores were entered at Block 2.
Both Block 1 and 2 displayed non-significant overall block effects, $\chi^2(1) = 1.94, p > .05$,
$-2LL = 179.16$, and $\chi^2(4) = 4.90, p > .05$, $-2LL = 176.09$, respectively (change from
previous block $\chi^2[3] = 3.07, p > .05$). However, when this process was repeated for IPV
recidivism, a significant Block 1 effect was revealed, $\chi^2(1) = 6.01, p = .014$, $-2LL =
83.86$. Upon the PCL-R, SARA, and SIR-R1 being entered within Block 2, the overall
block effect was diminished, $\chi^2(4) = 7.89, p = .096$, $-2LL = 81.82$. As with violent
recidivism, the PCL-R, SARA, and SIR-R1 failed to provide any significant incremental
validity relative to the ODARA in predicting IPV recidivism ($\chi^2[3] = 2.04, p > .05$).

Post hoc analyses for violent and IPV recidivism were conducted using the
DVRAG with the PCL-R score removed (i.e., weighted ODARA; ODARA_w) to examine
whether $F_1$ or $F_2$ of the PCL-R provided any incremental validity relative to the weighted
ODARA. To do so, a hierarchical Cox regression analysis was conducted with the ODARA entered at Block 1 and F₁ and F₂ entered at Block 2 using the backward Wald stepwise procedure. The backward stepwise procedure was selected to decrease the risk of a Type II error occurring due to suppressor effects (Field, 2009). In predicting violent recidivism, a significant Block 1 effect was observed for the ODARA, \( \chi^2 (1) = 4.81, p = .028, -2LL = 174.15 \). Similarly, when F₁ and F₂ entered in Step 1 of Block 2 there remained a significant overall block effect, \( \chi^2 (3) = 13.45, p = .004, -2LL = 166.95 \), which was a significant improvement over the previous block (\( \chi^2 [2] = 7.20, p = .027 \)). However, during Step 2 of Block 2, F₂ was removed from the equation leaving only the ODARA and F₁. The overall significant block effect for Block 2 at Step 2 was retained, \( \chi^2 (2) = 11.57, p = .003, -2LL = 168.11 \), and while the addition of F₁ displayed a significant improvement over the contribution provided by the ODARA alone, \( \chi^2 (1) = 6.05, p = .014 \), there were no statistically significant differences between Step 1 and Step 2 of Block 2 (\( \chi^2 [1] = 1.16, p > .05 \)).

In predicting IPV recidivism, there was slight variation in comparison to the previous analysis. Trend level significance was observed for the overall Block 1 effect for the ODARA, \( \chi^2 (1) = 3.74, p = .053, -2LL = 85.46 \). When F₁ and F₂ entered in Step 1 of Block 2, a significant overall block effect was observed, \( \chi^2 (3) = 9.34, p = .025, -2LL = 80.79 \). However, this was a non-significant improvement over the previous block (\( \chi^2 [2] = 4.67, p = .097 \)). Again during Step 2 of Block 2, F₂ was removed from the equation leaving only the ODARA and F₁. The overall significant block effect for Block 2 at Step 2 was retained, \( \chi^2 (2) = 8.37, p = .015, -2LL = 81.18 \), and while the addition of F₁

---

7Backward stepwise procedures were conducted using both the Wald statistic and the likelihood ratio as the Wald statistic may be unreliable (Field, 2009). However, both analyses yielded identical results within the current sample. Therefore, only the results using the backward Wald procedure are discussed.
displayed a significant improvement over the contribution provided by the ODARA \textsubscript{w} alone, $\chi^2 (1) = 4.28, p = .039$, there were no statistically significant differences between Step 1 and Step 2 of Block 2 ($\chi^2 [1] = 0.39, p > .05$). Overall, while \( F_1 \) provided significant incremental contributions relative to the ODARA \textsubscript{w} in predicting violent and IPV recidivism, it was not a statistically significant improvement over the contributions made by both \( F_1 \) and \( F_2 \).
Table 12
Hierarchical Cox Regression Survival Analysis for Violent and IPV Recidivism

<table>
<thead>
<tr>
<th>Cox regression analysis conducted</th>
<th>Violent recidivism</th>
<th>IPV recidivism&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>$SE$</td>
</tr>
<tr>
<td>Block 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ODARA</td>
<td>.172</td>
<td>.125</td>
</tr>
<tr>
<td>Block 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ODARA</td>
<td>.159</td>
<td>.139</td>
</tr>
<tr>
<td>PCL-R</td>
<td>.062</td>
<td>.041</td>
</tr>
<tr>
<td>SARA</td>
<td>-.001</td>
<td>.040</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>.007</td>
<td>.038</td>
</tr>
<tr>
<td>$N = 91$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Post hoc analyses

| Block 1                          |        |      |       |       |         |        |      |       |       |         |
| ODARA<sub>w</sub>               | .081   | .037 | 4.78* | 1.09  | [1.008, 1.166] | .104   | .054 | 3.71† | 1.11  | [0.998, 1.233] |
| Block 2                          |        |      |       |       |         |        |      |       |       |         |
| Step 1                           | ODARA<sub>w</sub> | .081 | .036 | 4.97* | 1.08  | [1.010, 1.164] | .106   | .052 | 4.10* | 1.11  | [1.003, 1.231] |
| PCL-R F<sub>1</sub>              | .171   | .063 | 7.31**| 1.19  | [1.048, 1.344] | .190   | .090 | 4.23* | 1.21  | [0.764, 1.149] |
| PCL-R F<sub>2</sub>              | -.076  | .071 | 1.15  | 0.93  | [0.806, 1.065] | -.065  | .104 | 0.39  | 0.94  | [0.764, 1.149] |
| $N = 88$                         |        |      |       |       |         |        |      |       |       |         |

Block 2                          |        |      |       |       |         |        |      |       |       |         |
| Step 2                           | ODARA<sub>w</sub> | .076 | .037 | 4.33* | 1.08  | [1.004, 1.160] | .100   | .051 | 3.82† | 1.11  | [1.000, 1.223] |
| PCL-R F<sub>1</sub>              | .154   | .061 | 6.33* | 1.17  | [1.035, 1.316] | .187   | .091 | 4.22* | 1.21  | [1.009, 1.441] |

Note. ODARA = Ontario Domestic Assault Risk Assessment; ODARA<sub>w</sub> = Weighted ODARA; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1.

<sup>a</sup>For the IPV recidivism analyses, 1 case was dropped due to being censored prior to the earliest event.

<sup>†</sup>$p < .10$, <sup>*</sup>$p < .05$, <sup>**</sup>$p < .01$. 
Discussion
Presently, while this is the second independent cross-validation of the ODARA and DVRAG, the current sample constitutes the highest risk sample to date in which the ODARA and DVRAG were validated. Representing a relatively homogenous group of high risk violent offenders, the offenders within the current study were referred to either the moderate or high intensity FVPP as a result of prior offending against a current or former marital, cohabiting, or dating partner. Among the sample, 81.9% had committed violence against an intimate partner prior to the index offence of which a small proportion of offenders (i.e., 10.6%) exhibited extreme forms of IPV with 3 cases resulting in the death of their intimate partner, whereas 7 resulted in hospitalization. Similarly, 81.9% of the sample had evidence of prior violence against a victim who was not their current or previous intimate partner or her children suggesting that only a small proportion of offenders were exclusively violent towards their intimate partners. While the sexual orientation of the sample was predominantly heterosexual (i.e., exhibited a history of female intimate partners only), 1 case was found in which the offender was bisexual and had violently offended against both male and female intimate partners. Moreover, the current sample may not necessarily be representative of the true federal offender population as only a single Aboriginal offender was included in the current study. However, this lack of Aboriginal offenders being included within the sample was the result of the SIR-R1 score being part of the inclusion criteria as the SIR-R1 is not administered to Aboriginal offenders as per CSC policy (Nafekh & Motiuk, 2002); although the single case within the current study appears to be an exception.

Although the previous independent cross-validation of the ODARA and DVRAG utilized the risk category scores within the predictive validity analyses (Rettenberger &
Eher, 2012), the current study examined the predictive validity of the total scores. Notwithstanding this minor difference, similar to the study conducted by Rettenberger and Eher, a number of important findings emerged within the current study. First, both the DVRAG and ODARA displayed high inter-rater reliability and predictive accuracy with respect to IPV recidivism. Similarly, the DVRAG was an efficacious predictor of any and violent recidivism, whereas the ODARA displayed significantly lower predictive accuracy for the latter when compared to the DVRAG. This appeared to be the result of combining the PCL-R with the weighted ODARA items. Accuracy of the SARA total score and subscales in predicting IPV recidivism ranged from slightly above chance levels to moderate levels of predictive accuracy (i.e., AUC = .508 to .635). Despite prior research finding moderate to high predictive validity for the SIR-R1 in predicting violent recidivism (e.g., Campbell et al., 2009), its performance within the current study was considered poor (AUC = .594). However, the SIR-R1’s poor predictive validity with respect to IPV recidivism was not surprising as prior research found it to be worse than chance (Gibas et al., 2008). Consistent with the general violence literature (e.g., Kroner & Mills, 2001; Kroner et al., 2005), no statistically significant differences were noted between the AUC values calculated for IPV recidivism; however, due to the low base rate for IPV recidivism, there may have been insufficient power to detect significant differences between the measures. Significant differences were noted, however, between the DVRAG and the SARA in predicting violent recidivism; although once adjusted for multiple pairwise comparisons the significance of these differences was lost. Moreover, treatment dosage did not appear to have a detrimental impact on the predictive validity of

---

*Predictive validity analyses excluded the risk category scores as 66.0% and 77.7% of the participants within the current study fell within the 7+ risk category for the DVRAG and ODARA, respectively, thus restricting the range of scores.*
either the DVRAG or ODARA further supporting the notion that actuarial risk assessment measures comprised of static or historical risk factors are less influenced by treatment effects. Indeed, only the predictive validity of the SARA appeared to be effected by the level of treatment dosage which may be related to its incorporation of dynamic risk variables. This result was also evident as no statistically significant mean differences were observed between the HIFVPP and MIFVPP on either the DVRAG or ODARA, whereas the two groups did differ in IPV risk as measured by the SARA and its various subscales. Additionally, the association between the SARA and both the DVRAG and ODARA was only considered moderate at best.

As noted by Mills and colleagues (2011), however, static risk factors such as those found within the DVRAG and ODARA do have a potential to change over the long-term; particularly in the event that the offender commits another violent offence. Although changes in offending behaviour may impact the total scores, they should not alter the actuarial risk estimate for the offender (Hilton et al., 2010). For example, if during an offender’s prior IPV offence he previously assaulted and threatened to kill during the assault an intimate partner who had more than one biological child from a previous partner, was pregnant at the time of the index assault, was confined during the index assault, and happened to have a prior history of drug abuse and provides care for more than one child under the age of 18, his score from these offence and victim characteristics alone would equal 16 points on the DVRAG. Now let us say, for arguments sake, the offender leaves his previous partner and becomes involved with a new intimate partner who has no children and no history of drug abuse and assaults her 8 months into their cohabiting relationship. However, during this assault he neither
confines the victim nor threatens to kill her and during a subsequent interview with the police she voices no concerns regarding the probability of another offence occurring. Rescoring of the DVRAG on this more recent hypothetical IPV index offence would result in a score of -3. Supposing that the offender scored at the highest values for the remaining 7 items at the time of his previous IPV offence, this would give him a score of 30 on those items. Now, using information from the previous IPV offence would yield a total score of 46 whereas using information from the most recent IPV offence would yield a total score of 27. While this represents a substantial difference of 19 points, visual inspection of the normative data presented by Hilton and colleagues (2010) reveals that neither score alters the actuarial estimates as both of these scores fall within the highest risk category for the DVRAG and are distributed approximately around the 98th and 99th percentile, respectively. Although this example is purely hypothetical, it does illustrate that at the clinical or applied level such an extreme fluctuation in IPV risk may not impact the overall risk estimate of the DVRAG (i.e., the offender would still be considered a high risk for IPV and the recidivism estimates would remain unchanged). Yet the fluctuations in IPV risk such as those illustrated by this example could have a detrimental impact at the statistical or empirical level with respect to the measures predictive accuracy. Theoretically, it seems somewhat counter-intuitive that an offender assessed as high risk can subsequently reduce his risk simply by committing another violent offence against an intimate partner. However, this may only be a reflection of how the dynamics and context of an intimate relationship will fluctuate between intimate partners. Regardless of these observations, the results of the current study suggest that the DVRAG is a significant predictor of IPV recidivism. Thus, items pertaining to the
context of the relationship combined with the behaviours during the index assault, while they may fluctuate between intimate partners, may be what are driving the predictive validity of the measure as opposed to the more static/historical items.

With respect to the other risk assessment measures, while it was not possible to analyze the inter-rater reliability of the SARA total or subscale scores, inter-rater reliability analyses for the PCL-R and SIR-R1 found high levels of agreement within an applied setting. While F1 of the PCL-R displayed the lowest ICC, it was among the strongest predictors of recidivism within the current study. This would suggest that, contrary to the argument proposed by Rufino and colleagues (2011), the poor predictive accuracy exhibited by F1 may go beyond measurement error and may relate to the type of violence being predicted. In addition, psychopathic personality as represented by F1 may be relating to IPV as those high on F1 could potentially be exhibiting increased traits associated with superficial charm, grandiosity, and manipulation. Such traits may assist the offender in acquiring new intimate partners at a faster rate thus granting them increased access to victims, in contrast to those low on such traits. Of note is that two items which do no load onto any factor relate to the presence of many short-term marital relationships and promiscuous sexual behaviour. Future research with IPV offenders could focus on such traits to determine if the frequency in which offenders engage in intimate relationships is significantly higher for those scoring high on F1 (particularly Facet 1), and, if so, does this relate to increased risk for IPV. Despite efforts to obtain hard copy PCL-R scores, total and factor scores had to be derived using percentile rankings for the majority of the sample. This limited the analyses as Facet scores could not be obtained for the entire sample. Moreover, while the current study found strong
predictive accuracy for $F_1$, $F_2$ scores were slightly biased as for 17 cases the percentile rankings were those of the 1991 version of the PCL-R. Of note is that this earlier version did not include Item 20 (Criminal Versatility) within its factor structure (i.e., $F_2$).

However, among the 77 cases with $F_2$ derived from the 2003 version of the PCL-R, the association with IPV recidivism was still non-existent ($r_{pb} = .02$, AUC = .509).

Nonetheless, future research should focus on the relative contributions of the various PCL-R factor and facet scores to determine whether supplementing $F_1$ or Facets 1 or 2 in place of the PCL-R total score would optimize predictive validity of the DVRAG.

Although the results of the current study are promising, there were a number of limitations specific to the DVRAG and ODARA. First, while the two measures were considered robust predictors of recidivism despite missing items and potential treatment effects, there was a noticeable impact with respect to the structural reliability analysis due to the number of missing items, particularly the item relating to victim concern. The poor corrected item-total correlations found for some of the items may be a reflection of the item selection process used during the construction of the ODARA, as items selected and retained based on their associations with outcome may not necessarily be related to each other, or may reflect an underlying multi-dimensional structure. Although the latter should not affect the value of Cronbach’s alpha as multi-dimensional measures can produce large alpha values (see Schmitt, 1996). As such, the results of the structural reliability analyses must be interpreted with caution. Although retrospectively scored based on file review alone, the content contained within the federal correctional files was considered extensive and, at times, exhaustive. However, for some of the cases, detailed information pertaining to prior offences or provincial sentences was not consistently
available. Similar to the study conducted by Hilton, Harris, Popham and colleagues (2010), victim interviews were rare. However, unlike their study, identification of prior domestic assaults was readily available through the family violence assessments which utilized information stemming from criminal records, police reports, and self-report from the offender or a family member. Information obtained from self-reports was considered integral to the coding procedure for a small number of cases as the historical police reports had been expunged by the investigating police department prior to the offender entering federal custody.

An argument can be made that conducting psychological risk assessments at the federal level represents a top-down procedure in which the scoring of the measures is based on a semi-structured interview conducted with the offender in conjunction with a thorough review of the offender’s correctional files and criminal history. In contrast, the items within the ODARA were designed and selected using a bottom-up procedure in which scoring was based on semi-structured interview conducted with the victim and a brief review of the offender’s criminal history. Whereas specific victim and index related information may be readily available to frontline police personnel, this is not necessarily the case for mental health professionals at the federal level.

Whereas mental health professionals scoring the DVRAG or ODARA could potentially discern missing item information through interviews conducted with the victim or offender during the intake assessment process, this could result in unnecessary expenditure of resources and may prove fruitless as only a small proportion of the current sample (i.e., 14.9% or $n = 14$) retained their intimate relationship with the victim of the index offence at the time of admission to federal custody. Furthermore, comparing those
offenders with and without an IPV related federal index offence revealed that the
offenders with an IPV related federal index offence had significantly shorter periods of
time between the IPV index offence date and sentence commencement date in
comparison to those who did not ($t(92) = 6.07$, $p < .001$, $M = 11.55$ vs. $40.66$ months,
respectively). This suggests that the likelihood of conducting a victim interview would be
low as the average amount of time elapsed between the IPV index offence and admission
to federal custody for over half the sample was, on average, over 3 years. However, for
the majority of offenders with an IPV related index federal offence, victim impact
statements were readily available through OMS of which a number of them included a
statement made by the victim pertaining to the offender’s potential for future violence
towards them. A caveat, however, is that three offenders murdered their intimate partner
thus rendering victim concern unattainable. Nonetheless, prorating the DVRAG in the
event of missing information does not appear to impact its predictive validity.

Second, coding of the DVRAG and ODARA items was based on retrospective
file review in which being blind to the SARA or SIR-R1 scores and recidivism rates was
not feasible. While this may have had a biasing effect on the predictive validity of the
two measures, such effects were tempered by the high inter-rater reliability between the
current author’s scores and those of an independent rater blind to the recidivism rates,
and by the relative static nature of the items and their focus on prior offending behaviour
and behaviour exhibited during the IPV index offence. As noted within the sample
characteristics, less than half (i.e., 40.4%) of the sample were serving their sentence due
to an IPV related index offence. Therefore, a number of offenders had already reoffended
violently since their IPV index offence. Closer inspection revealed that of the 56
offenders who did not enter federal custody due to an IPV related index offence, the majority had been incarcerated due to other forms of violent offending against non-intimate partners (i.e., robbery = 22, theft and/or break and enter = 8, assault related offences = 7 [e.g., assault, assault causing bodily harm, etc.], drug related offences = 7, driving related offences = 5, manslaughter = 3, weapons offences = 2, arson =1, and attempted murder = 1). However, as outlined within the ODARA and DVRAG coding procedures, events such as these are not coded as they did not occur prior to the IPV index related offence. For example, 4 cases were rated as having no prior nondomestic incidents despite serving their current federal sentence for a violent offence. Therefore, the DVRAG and ODARA were actually at a disadvantage in comparison to the PCL-R, SARA, and SIR-R1 as these latter measures did account for this information.

Similar to the observations made by Hilton, Harris, Popham and colleagues (2010), coding of IPV recidivism was problematic for a total of 5 cases within the current sample as the relationship between the victim and offender could not be ascertained. This issue notwithstanding, both the DVRAG and ODARA produced large AUC values suggesting that, contrary to results found by Hilton and Harris (2009), ambiguity did not have a detrimental impact on the predictive validity of the two measures. In fact, excluding the ambiguous cases decreased predictive validity yielding AUC values .685 and .608 for the DVRAG and ODARA, respectively. Also, the inclusion of IPV recidivism in the definition of violent recidivism may have artificially inflated the AUC values for the DVRAG and ODARA. Of note is that the base rate within the current study for IPV recidivism (i.e., 12.8%) appeared quite low in comparison to those found in previous studies (e.g., 27.3% for Hilton, Harris, Popham et al., 2010; 21.2% for
Rettenberger & Eher, 2012). However, this comparatively low base rate may be a function of the relatively short follow-up period of 3.6 years as the follow-up periods for the two studies previously reported were approximately 4.58 and 7.98 years, respectively. Moreover, the current study did not account for the length of time in which the offenders had been under federal supervision. It is not uncommon for offenders with a history of IPV to have special conditions imposed by the PBC wherein their intimate relationships must be reported and monitored accordingly. Therefore, a longer follow-up period would have increased the likelihood that all of the offenders had surpassed their WED resulting in the subsequent expiration of their parole conditions. Together, these factors may have impacted the observed recidivism rates. Indeed, 10 cases within the current study were still serving their federal sentence within the community under federal supervision at the time of follow-up.

Another potential limitation relates to the statistical analyses employed within the current study. Although statistics and statistical techniques such as AUCs, correlation coefficients, Cohen’s $d$, and Cox regression, have become the standard for examining the predictive validity of violence risk assessment measures, they are nonetheless limited in their applications. For example, correlation coefficients can be heavily influenced by fluctuations in recidivism base rates (McGrath & Meyer, 2006), whereas AUCs can only be applied to dichotomous outcome variables and are unable to compensate for variations in follow-up time (Harris & Rice, 2003). Similarly, while Cox regression controls for variations in follow-up time, it too can only be applied to a dichotomous outcome variable. However, statistical techniques have been developed to calculate three-way ROC curves for trichotomous outcome variables (Mossman, 1999) and time-dependent
ROC curves (e.g., Chambless & Diao, 2006; Heagerty, Lumley, & Pepe, 2000). Application of such techniques may overcome many of the limitations of the statistical techniques commonly used for recidivism studies. However, despite being in existence for more than a decade, such statistical analyses have yet to be applied to the prediction of violent recidivism. Recent advancements to the assessment of violence risk have unveiled other pertinent information that can impact predictive accuracy such as release context (Kroner et al., 2011) and pre-post treatment change (Beggs & Grace, 2010). Future research with IPV offenders should address these limitations by adopting these more advanced statistical techniques and assessment procedures.

Revisiting the Meta-Analytic Investigation and Moderator Analysis

For descriptive purposes the meta-analytic results reported earlier were updated to include the results of the present study. Utilizing the same statistical techniques described within the introduction, the updated meta-analytic results for the five risk assessment measures in predicting IPV are presented in Table 13. Updated average base rates for IPV recidivism were as follows: DVRAG = 29%, k = 5; ODARA = 30%, k = 8; PCL-R = 25%, k = 4; SARA = 32%, k = 10; and the SIR-R1 = 20%, k = 2. As before, the DVRAG yielded the highest aggregated AUC value in comparison to the remaining risk assessment measures (AUC$^w = .710$, 95% CI$^w [.673, .746]$). Again, no significant heterogeneity among the weighted AUC values was noted. Indeed, very little variation was noted between the previous and updated meta-analytic results; with the exception of slight reductions in the width of the 95% confidence intervals. A slight increase in the weighted AUC was noted for the SIR-R1; however, the final value did not surpass chance
ASSESSING RISK FOR INTIMATE PARTNER VIOLENCE

levels. This lends increasing support for the use of specialized risk assessment measures such as the DVRAG and ODARA when assessing risk in IPV offenders.

Table 13

Weighted AUC Values for Five Risk Assessment Measures in Predicting IPV

<table>
<thead>
<tr>
<th>Measure</th>
<th>n (k)</th>
<th>AUC&lt;sub&gt;w&lt;/sub&gt;</th>
<th>95% CI&lt;sub&gt;w&lt;/sub&gt;</th>
<th>z</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVRAG</td>
<td>873 (5)</td>
<td>.710</td>
<td>[.673, .746]</td>
<td>11.32*</td>
<td>0.73</td>
</tr>
<tr>
<td>ODARA</td>
<td>1,517 (8)</td>
<td>.679</td>
<td>[.650, .709]</td>
<td>11.99*</td>
<td>7.12</td>
</tr>
<tr>
<td>PCL-R</td>
<td>893 (4)</td>
<td>.655</td>
<td>[.617, .693]</td>
<td>7.91*</td>
<td>2.08</td>
</tr>
<tr>
<td>Factor 1</td>
<td>889 (4)</td>
<td>.583</td>
<td>[.543, .623]</td>
<td>4.04*</td>
<td>7.14</td>
</tr>
<tr>
<td>Factor 2</td>
<td>890 (4)</td>
<td>.664</td>
<td>[.626, .703]</td>
<td>8.37*</td>
<td>2.63</td>
</tr>
<tr>
<td>SARA</td>
<td>2,696 (10)</td>
<td>.633</td>
<td>[.609, .656]</td>
<td>10.93*</td>
<td>14.75</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>202 (2)</td>
<td>.501</td>
<td>[.400, .602]</td>
<td>0.02</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Note. n = total sample size; k = number of effect sizes; AUC<sub>w</sub> = mean weighted effect size; CI<sub>w</sub> = 95% confidence interval of AUC<sub>w</sub>; z = significance test of AUC<sub>w</sub>; Q = test of homogeneity. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1.

* p < .001.

In addition, AUC values were collected for F<sub>1</sub> and F<sub>2</sub> from studies that have examined the PCL-R in predicting IPV recidivism (Grann & Wedin, 2002; Hilton et al., 2008a; Rettenberger & Eher, 2012a). While both the current study and the study by Grann and Wedin found high predictive accuracy for F<sub>1</sub> (AUC = .685 and .700, respectively), the remaining two studies conducted by Hilton and colleagues and Rettenberger and Eher found F<sub>1</sub> to be a poor predictor of IPV recidivism (AUC = .578 and .424, respectively). These discrepancies notwithstanding, the aggregated results found F<sub>1</sub> to be a modestly significant predictor of IPV recidivism (AUC<sub>w</sub> = .583, p <
.001), whereas $F_2$ was within the moderate range and was highly significant ($AUC_w = .664, p < .001$).

With the exception of slight variations noted for the updated $Q_A$ and $p$-values, no changes were observed with respect to the significant differences previously noted in the predictive accuracy between the DVRAG and the PCL-R, DVRAG and the SARA, and between the ODARA and the SARA. Once again, after applying the Bonferroni adjustment for the multiple pairwise comparisons, significant differences remained only between the DVRAG and the SARA. As before, the $AUC_{between}$ calculated for the DVRAG and SARA again achieved significance ($p = .012$), however, trend level significance was observed for the $AUC_{between}$ calculated for the DVRAG and PCL-R ($p = .050$) and ODARA and SARA ($p = .056$).

Table 14

<table>
<thead>
<tr>
<th>Pairwise comparison (k)</th>
<th>$AUC_{between}$</th>
<th>$Q_A$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVRAG versus ODARA (13)</td>
<td>.68</td>
<td>1.52</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>DVRAG versus PCL-R (9)</td>
<td>.90</td>
<td>3.71</td>
<td>1</td>
<td>.054</td>
</tr>
<tr>
<td>DVRAG versus SARA (15)</td>
<td>.91</td>
<td>11.65</td>
<td>1</td>
<td>.0006</td>
</tr>
<tr>
<td>ODARA versus PCL-R (12)</td>
<td>.77</td>
<td>0.69</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>ODARA versus SARA (18)</td>
<td>.77</td>
<td>5.73</td>
<td>1</td>
<td>.017</td>
</tr>
<tr>
<td>PCL-R versus SARA (14)</td>
<td>.50</td>
<td>0.90</td>
<td>1</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

*Note. $k =$ number of effect sizes per pairwise comparison; $Q_A =$ between-level $Q$. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment. Bonferroni adjustment: $p < .05 = .008$, $p < .01 = .002$, $p < .001 = .0002$. 
Conclusions

In summary, the results of the current study in conjunction with the results of the meta-analysis suggest that the DVRAG and ODARA are the most efficacious predictors of IPV recidivism relative to the PCL-R, SARA, and SIR-R1. While non-significant differences were noted at the study level between the five risk assessment measures in predicting IPV recidivism, at the aggregate level the results indicate the DVRAG significantly outperforms the SARA even after accounting for multiple pairwise comparisons. Moreover, the poor predictive accuracy of the general criminal risk variables (i.e., SIR-R1 total score, PCL-R F2 score, and the SARA Criminal History subscale score) provide further empirical support for the use of specialized risk assessment measures with IPV offenders. Although within the current study the SARA was considered to be only a modest predictor of IPV recidivism, accounting for treatment dosage did improve its association with IPV recidivism outcomes. Overall, while the AUC values within the current study ranged from modest to large (Rice & Harris, 2005), these values are representative of a level of predictive accuracy that is “far from ideal” (Hanson, 2009, p. 177). Therefore, it is only through the adoption and application of increasingly advanced statistical analyses (e.g., time-dependent ROCs) and assessment procedures (e.g., dynamic-actuarial risk assessment; Mills et al., 2011) that the field of violence risk assessment will advance. Undoubtedly, advancements such as these will have far reaching implications with respect to public safety and offender liberty.
References


perspectives (pp. 314-350). Toronto, Ontario, Canada: University of Toronto Press.


Appendix A: AUC Meta-Analysis Syntax

matrix.

* input variables for the observed auc values = auc, total sample size = n, Wilcoxon standard error = se_auc, and the average base rate (expressed in decimal form) = base.
get auc /variables = auc.
get n /variables = n.
get se_auc /variables = se_auc.
get base /variables = base.

* get number of studies (k) from number of rows.
compute k = nrow(auc).
compute w = 1/((se_auc)***(se_auc)).

* sum weights, and compute total sample size, and the overall average base rate.
compute sw = csum(w).
compute sample = csum(n).
compute ave_base = csum(base)/k.

print /title = "********** META-ANALYSIS OF AUC VALUES **********".
print (k, sample}/clabels=k, sample/format=f9.0/title = "NUMBER OF STUDIES and NUMBER OF PARTICIPANTS".
print {ave_base}/clabels=baserate/format=f9.2/title= "AVERAGE BASE RATE".

* obtain weighted means.
compute mean = {t(w)*auc/sw}.

* calculate standard errors and confidence intervals for means, and Q-statistic.
compute sem = {sqrt(1/sw)}.
compute ci = {mean - 1.96*sem, mean + 1.96*sem}.
compute chi = csum((w)***(auc-mean)**2).
compute chisig = 1 - chicdf(chi, k-1).

compute z = (mean - 0.5)/sem.
compute pz = {2*(1 - cdftnorm(z))}.

print /title = "********** FIXED-EFFECTS MODEL **********".
print {mean(1),sem, ci(1,1), ci(1,2), z(1), pz(1)}/clabels='Mean AUC', 'Weighted SE', 'Lower CI', 'Upper CI', 'z', 'p'/format=f9.3/title='MEAN AUC, WEIGHTED STANDARD ERROR, 95% CONFIDENCE INTERVALS, and Z-TEST'.
print {chi, k-1, chisig}/clabels=Chi2, df,p/format=f9.3/title = "HOMOGENEITY TEST: Q STATISTIC (Goodness of Fit)".
end matrix.
Example of SPSS output:

**Matrix - ODARA**

[DataSet4] E:\SPSS Meta-analysis Syntax\roc

Example of SPSS output:

Matrix - ODARA

[DataSet4] E:\SPSS Meta-analysis Syntax\roc
curves_meta_moderator2012.sav

Run MATRIX procedure:

********** META-ANALYSIS OF AUC VALUES **********

NUMBER OF STUDIES and NUMBER OF PARTICIPANTS

<table>
<thead>
<tr>
<th>k</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1423</td>
</tr>
</tbody>
</table>

AVERAGE BASE RATE

<table>
<thead>
<tr>
<th>baserate</th>
</tr>
</thead>
<tbody>
<tr>
<td>.33</td>
</tr>
</tbody>
</table>

********** FIXED-EFFECTS MODEL **********

MEAN AUC, WEIGHTED STANDARD ERROR, 95% CONFIDENCE INTERVALS, and Z-TEST

<table>
<thead>
<tr>
<th>Mean AUC</th>
<th>Weighted</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>.678</td>
<td>.015</td>
<td>.648</td>
<td>.708</td>
<td>11.642</td>
<td>.000</td>
</tr>
</tbody>
</table>

HOMOGENEITY TEST: Q STATISTIC (Goodness of Fit)

<table>
<thead>
<tr>
<th>Chi2</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.915</td>
<td>6.000</td>
<td>.329</td>
</tr>
</tbody>
</table>

------- END MATRIX -------
Appendix B: AUC Power Analysis Syntax

* input variables for the estimated auc value = auc, the null auc value (i.e., .50) = null, desired alpha level (expressed as a z-score; e.g., 1.645 for a one-tailed test at .05) = z_alpha, desired power (expressed as a z-score; e.g., .840 for 80% power) = z_beta, and the hypothetical base rate = br.

get auc /variables = auc.
get null /variables = null.
get z_alpha /variables = z_alpha.
get z_beta /variables = z_beta.
get br /variables = br.

compute dif = auc - null.
compute Q1_auc = auc / (2 - auc).
compute Q2_auc = (2 * (auc * auc)) / (auc + 1).
compute Q1_null = null / (2 - null).
compute Q2_null = (2 * (null * null)) / (null + 1).
compute V1 = Q1_null + Q2_null - (2 * (null * null)).
compute V2 = Q1_auc + Q2_auc - (2 * (auc * auc)).
compute d = (z_alpha * (SQRT(2 * V1))) + (z_beta * (SQRT(V1 + V2))).
compute n = (d / dif) * (d / dif).
compute rndn = RND(n).
compute nt = rndn * 2.
compute n_rec = RND((br * nt)).
compute n_non = (nt - n_rec).
compute a = (auc * (1 - auc)) + ((n_rec - 1) * (Q1_auc - (auc * auc))) + ((n_non - 1) * (Q2_auc - (auc * auc))).
compute b = n_rec * n_non.
compute se_w = SQRT(a / b).
compute ci = {auc - 1.96 * se_w, auc + 1.96 * se_w}.

compute z = (auc - 0.5) / se_w.
compute pz = 2 * (1 - cdfnorm(z)).

print /title '***** ROC POWER ANALYSIS *****'.
print {auc} /clabels=AUC/format=f9.3 /title = "Estimated AUC value".
print {z_alpha, z_beta} /clabels=alpha, beta/format=f9.3 /title = "Probability of Type I and Type II error".
print {n} /clabels=n/format=f9.2 /title = "Number of required participants per group (assuming 50% base rate)".

print {nt} /clabels=N/format=f9.0 /title = "Total number of participants required for study".
compute BR = br *100.
print {BR (1), n_rec(1), n_non(1)}/clabels='baserate', '# of rec', '# of non-rec'/format=f9.0/title = "Estimated base rate (%) with number of recidivists and non-recidivists".
print {auc(l), ci(l,1), ci(l,2), z(l), pz(l), se_w}/clabels='AUC', 'Lower CI', 'Upper CI', z, p, 'Wilcoxon se'/format=f15.3/title='AUC, lower and upper 95% confidence intervals, z-test, and calculated Wilcoxon standard error'.
end matrix.
Example of SPSS output:

Run MATRIX procedure:

***** ROC power analysis *****

Estimated AUC value

<table>
<thead>
<tr>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>.709</td>
</tr>
</tbody>
</table>

Probability of Type I and Type II error

<table>
<thead>
<tr>
<th>alpha</th>
<th>beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.645</td>
<td>.840</td>
</tr>
</tbody>
</table>

Number of required participants per group (assuming 50% base rate)

<table>
<thead>
<tr>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>45.44</td>
</tr>
</tbody>
</table>

Total number of participants required for study

<table>
<thead>
<tr>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
</tr>
</tbody>
</table>

Estimated base rate (%) with number of recidivists and non-recidivists

<table>
<thead>
<tr>
<th>baserate</th>
<th># of rec</th>
<th># of non</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>18</td>
<td>72</td>
</tr>
</tbody>
</table>

AUC, lower and upper 95% confidence intervals, z-test, and calculated Wilcoxon standard error

<table>
<thead>
<tr>
<th>AUC</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>z</th>
<th>p</th>
<th>Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>.709</td>
<td>.564</td>
<td>.854</td>
<td>2.829</td>
<td>.005</td>
<td>.074</td>
</tr>
</tbody>
</table>

------- END MATRIX ------
Appendix C: Certificate of Ethics Approval

Carleton University
Psychology Research Ethics Board

Certificate of Ethics Clearance

<table>
<thead>
<tr>
<th>Principal Investigator</th>
<th>Department</th>
<th>Study Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew Gray</td>
<td>Psychology</td>
<td>11-139</td>
</tr>
</tbody>
</table>

Institution(s) where research will be conducted:
Regional Headquarters, Correctional Service Canada

Co-Investigators and other researchers:

<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>Jeremy Mills</td>
<td>Faculty</td>
</tr>
<tr>
<td>Researcher</td>
<td>Adelle Forth</td>
<td>Faculty</td>
</tr>
<tr>
<td>Other Research Personnel</td>
<td></td>
<td>Faculty</td>
</tr>
<tr>
<td>Faculty Sponsor</td>
<td></td>
<td>Faculty</td>
</tr>
</tbody>
</table>

Study Title: Assessing Risk among Domestically Violent Offenders: An Evaluation of the Psychometric Properties of the ODARA and DVRAG within a Sample of Incarcerated

Approval Date: 02/21/2012
Validity Term: Until Aug 31st Next

Comments:

Certification

The protocol describing the above-named project has been reviewed by Carleton University Psychology Research Ethics Board and the research procedures were found to be acceptable on ethical grounds for research involving human participants.

Chair, Ethics Committee for Psychology Research

This Certificate of Clearance is valid for the above term provided there is no change in the research procedures.

Close

Print
D: Research Application and Undertaking

**RESEARCH APPLICATION AND UNDERTAKING**

**Research Branch**

**NOTE - SEE REVERSE FOR TERMS AND CONDITIONS**

Date submitted: November 16, 2011 (RDC and RRC recommended)

**RESEARCH PROJECT - PROJET DE RECHERCHE**

**PROJECT TITLE - TITRE DU PROJET**

Assessing Risk among Violent Offenders: An Evaluation of the Psychometric Properties of the ODARA and DVRAG within a Sample of Incarcerated Offenders

**PROJECT DESCRIPTION - DESCRIPTION DU PROJET**

Purpose - But: The purpose of the study is to validate two recently published actuarial risk assessment measures designed for domestically violent offenders: the Ontario Domestic Assault Risk Assessment (ODARA) and the Domestic Violence Risk Appraisal Guide (DVRAG) and to comparatively evaluate the predictive validity of the ODARA, DVRAG, PCL-R, SIR-R1 and SARA in predicting recidivism.

Participants - Participants: The proposed study will consist of retrospectively coding file information from a sub-sample of 586 male federal offenders who met the criteria for participation in either the Moderate or High Intensity Family Violence Prevention Program. The offender sample will be the same one which was previously approved for use in pre-post efficacy of the Family Violence Prevention Program (FVPP) by Angela Connors and Dr. Jeremy Mills.

**TYPE/CLASS OF INFORMATION REQUESTED - TYPE/CATEGORIE DES RENSEIGNEMENTS DEMANDES**

Voluntary - Informed consent to participate - Consentement volontaire et éclairé pour participer

Criteria - Critères:

a. conformity with the principles of CCRA - conformité aux principes de la LCSM

b. contribution to the achievement of the Mission and the priorities of CSC - contribution à la réalisation de la Mission et aux priorités du Service correctionnel du Canada:

c. compliance with the Tri-councils Policy Statement on Ethical Conduct for Research Involving Humans - conformité à l'enquête de politique des trois Conseils intitulé Éthique de la recherche avec des êtres humains

d. level of disruption to the implementation of correctional objectives from an operational perspective - mesure dans laquelle le projet nuirait à la mise en œuvre des objectifs correctionnels sur le plan opérationnel

e. quality of the methodology - qualité de la méthode

f. qualifications of the researchers - compétences des chercheurs

g. anticipated benefit to corrections - avantages éventuels pour les services correctionnels

h. value for money - rapport valeur-coût

**PRIMARY RESEARCHER - CHERCHEUR PRINCIPAL**

Andrew Gray, B.A.H., B.S.T.
Graduate Student, Carleton University

Operational unit - Unité opér.
CSC

Sector - Secteur
Regional Headquarters

Region - Région
Ontario

Telephone number - N° de téléphone

**OTHER RESEARCHERS - AUTRES CHERCHEURS**

a) Name and title - Nom et titre

b) Name and title - Nom et titre

c) Name and title - Nom et titre

**Director General Research - Directeur général de la Recherche**

Signature

Brian A. Grant

**DATE**

Y-M-D

**DISTRIBUTION**

Original = Copy =

CSC/SNC: 954 (R-99-04)

7530-21-
ASSESSING RISK FOR INTIMATE PARTNER VIOLENCE

1

The provision of data for research purposes is approved by the Director General of Research. All participants shall adhere to the terms and conditions of the Research Application and Undertaking Form. The Service has the right to examine any results or reports in order to verify that no personal information has been compromised. No further personal information will be provided to the participant(s) if this agreement is violated.

Non-disclosure of personal information
The researcher or organization

I) agrees that information contained in or resulting from the data provided shall be rendered anonymous (i.e., coded and stripped of personal identifiers) as soon as possible

II) shall not contact the study’s participant(s) unless specifically authorized to do so and, if authorized, shall guarantee that any additional information obtained from these individuals will receive the same level of confidentiality as maintained for the original data

III) agrees that no subsequent disclosure of the information in a form that could reasonably be expected to identify the individual(s) to whom it relates will be made to any other person or third party not participating in the research project

Data location, consultation and disposal
For any data not secured within the physical confines of the Correctional Service of Canada, the participant(s) shall provide

I) a written statement specifying where the data or copies of records will be consulted or used

II) a list of any data linkages that have been approved and an undertaking not to perform other matches without written authorization

III) a written statement regarding the conditions governing disposal of identifiable personal data and, if applicable, the archival arrangements for working data and the relevant programs, codes and guides

Participant(s) acknowledgements
The participant(s)

I) recognizes that permission to conduct research may be withdrawn at any time for violations of standing orders, rules and regulations related to security or for unapproved deletions from the original proposal, or it may be temporarily suspended for operational reasons

II) will abide by the standing orders and rules of the institution or parole office, including those designed to ensure the researcher’s own safety

III) acknowledges that, other than the report submitted for publication, no further release of data shall be made without the permission of the Service and/or the Department

IV) will make appropriate acknowledgement in all subsequent reports of the Correctional Service of Canada’s sponsorship of the research, as well as in an appropriate disclaimer that the opinions and conclusions do not necessarily represent those of the Service and/or Department

V) acknowledges that ownership of the data rests with the Correctional Service of Canada

Approval
It is agreed that all parties signing this agreement and any attachments thereto are satisfied with and will adhere to the terms and conditions.
Appendix E: ODARA Certification

Andrew L. Gray
is registered by the
Research Department as a
Certified User of the
Ontario Domestic Assault
Risk Assessment

Certificate No. 12–083

May 22, 2012

N. Zoe Hilton, PhD.
Senior Research Scientist

Waypoint
CENTRE FOR MENTAL HEALTH CARE
Appendix F: Comparisons between the Moderate and High Intensity FVPP

<table>
<thead>
<tr>
<th>Measure</th>
<th>Moderate M (SD)</th>
<th>High M (SD)</th>
<th>F (df = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criminal history</td>
<td>3.85 (1.61)</td>
<td>3.76 (1.82)</td>
<td>0.07</td>
</tr>
<tr>
<td>Psychosocial adjustment</td>
<td>5.48 (2.32)</td>
<td>6.48 (3.03)</td>
<td>3.25†</td>
</tr>
<tr>
<td>Spousal assault history</td>
<td>6.25 (2.96)</td>
<td>8.64 (3.26)</td>
<td>13.02**</td>
</tr>
<tr>
<td>Alleged/most recent offense</td>
<td>0.95 (1.56)</td>
<td>2.30 (2.05)</td>
<td>12.78**</td>
</tr>
<tr>
<td>Other considerations</td>
<td>0.33 (1.04)</td>
<td>0.67 (1.47)</td>
<td>1.68</td>
</tr>
<tr>
<td>SIR-R1</td>
<td>6.69 (6.40)</td>
<td>2.82 (8.79)</td>
<td>5.99*</td>
</tr>
</tbody>
</table>

*Note. M = mean; SD = standard deviation; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1.
†p < .10, *p < .05, **p < .01

The Multivariate Analysis of Variance (MANOVA) comparing the MIFVPP (n = 61) and HIFVPP (n = 33) on the SARA subscales and SIR-R1 total score revealed significant differences on all but the Criminal History and Other Considerations subscales. Hence, those referred to the HIVPP were at significantly greater IPV risk as measured by the SARA, yet were at significantly lower general criminal risk as measured by the SIR-R1.
### Appendix G: Correlational Analysis between the PCL-R, SARA, and SIR-R1

<table>
<thead>
<tr>
<th>Risk Measure</th>
<th>n</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PCL-R</td>
<td>91</td>
<td>.75***</td>
<td>.77***</td>
<td>.26*</td>
<td>.17</td>
<td>.21*</td>
<td>.20†</td>
<td>.03</td>
<td>-.15</td>
<td>.42***</td>
</tr>
<tr>
<td>2. Factor 1</td>
<td>88</td>
<td>-</td>
<td>.24*</td>
<td>.10</td>
<td>-.05</td>
<td>-.02</td>
<td>.17</td>
<td>.09</td>
<td>-.07</td>
<td>.09</td>
</tr>
<tr>
<td>3. Factor 2</td>
<td>89</td>
<td>-</td>
<td>.30**</td>
<td>.27**</td>
<td>.39**</td>
<td>.15</td>
<td>-.08</td>
<td>-.18†</td>
<td>.53***</td>
<td></td>
</tr>
<tr>
<td>4. SARA</td>
<td>94</td>
<td>-</td>
<td>.36***</td>
<td>.67***</td>
<td>.81***</td>
<td>.50***</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Criminal history</td>
<td>94</td>
<td>-</td>
<td>.06</td>
<td>.14</td>
<td>-.08</td>
<td>-.22*</td>
<td>.38***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Psychosocial adjustment</td>
<td>94</td>
<td>-</td>
<td></td>
<td>.30**</td>
<td>.15</td>
<td>.04</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Spousal assault history</td>
<td>94</td>
<td>-</td>
<td></td>
<td>.26*</td>
<td>.16</td>
<td></td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Alleged/most recent offense</td>
<td>94</td>
<td>-</td>
<td></td>
<td>.09</td>
<td></td>
<td>-.25*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Other considerations</td>
<td>94</td>
<td>-</td>
<td></td>
<td></td>
<td>-.29**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. SIR-R1</td>
<td>94</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1.

†*p < .10, *p < .05, ** p < .01, *** p < .001.
Appendix H: Accuracy of Five Risk Assessment Measures in Predicting Revocation of Conditional Release ($N = 88$)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Revocation of Conditional Release (55.7%)</th>
<th>$r_{pb}$</th>
<th>AUC</th>
<th>95% CI&lt;sub&gt;NP&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVRAG</td>
<td></td>
<td>.08</td>
<td>.517</td>
<td>[.394, .640]</td>
</tr>
<tr>
<td>ODARA</td>
<td></td>
<td>-.00</td>
<td>.473</td>
<td>[.349, .598]</td>
</tr>
<tr>
<td>PCL-R&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>.15</td>
<td>.588</td>
<td>[.466, .711]</td>
</tr>
<tr>
<td>Factor 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>.15</td>
<td>.597</td>
<td>[.474, .721]</td>
</tr>
<tr>
<td>Factor 2&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>.13</td>
<td>.574</td>
<td>[.449, .699]</td>
</tr>
<tr>
<td>SARA</td>
<td></td>
<td>.03</td>
<td>.506</td>
<td>[.383, .629]</td>
</tr>
<tr>
<td>Criminal history</td>
<td></td>
<td>.02</td>
<td>.513</td>
<td>[.390, .635]</td>
</tr>
<tr>
<td>Psychosocial adjustment</td>
<td></td>
<td>-.09</td>
<td>.451</td>
<td>[.328, .575]</td>
</tr>
<tr>
<td>Spousal assault history</td>
<td></td>
<td>.17</td>
<td>.581</td>
<td>[.461, .701]</td>
</tr>
<tr>
<td>Alleged/most recent offense</td>
<td></td>
<td>-.10</td>
<td>.445</td>
<td>[.323, .567]</td>
</tr>
<tr>
<td>Other considerations</td>
<td></td>
<td>-.06</td>
<td>.462</td>
<td>[.339, .584]</td>
</tr>
<tr>
<td>SIR-R1</td>
<td></td>
<td>.12</td>
<td>.561</td>
<td>[.441, .682]</td>
</tr>
</tbody>
</table>

Note. DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist-Revised; SARA = Spousal Assault Risk Assessment; SIR-R1 = Statistical Information on Recidivism-Revised 1. AUC = Area under the curve of the receiver operating characteristic (ROC); 95% CI<sub>NP</sub> = Nonparametric 95% confidence interval of the AUC based on standard error derived from the Hanley and McNeil (1982) method calculated using SPSS. Base rate for revocation of conditional release is in parentheses.

<sup>a</sup> $N = 85$. Base rate for revocation of conditional release = 55.3%.

<sup>b</sup> $N = 82$. Base rate for revocation of conditional release = 57.3%.

<sup>c</sup> $N = 83$. Base rate for revocation of conditional release = 56.6%.