Risk Assessment in Intimate Partner Violence: A Systematic Review of Contemporary Approaches

Tonia L. Nicholls, PhD
University of British Columbia, Vancouver, British Columbia, Canada
Forensic Psychiatric Services Commission, British Columbia Mental Health & Addiction Services, Port Coquitlam, British Columbia, Canada
Simon Fraser University, Burnaby, British Columbia, Canada

Michelle M. Pritchard, BSc
Forensic Psychiatric Services Commission, British Columbia Mental Health & Addiction Services, Port Coquitlam, British Columbia, Canada
Simon Fraser University, Burnaby, British Columbia, Canada

Kim A. Reeves, MA
Simon Fraser University, Burnaby, British Columbia, Canada

Edward Hilterman, MSc
Tilburg University, Tranzo—Scientific Center for Care and Welfare, The Netherlands
Justa Mesura, Consultancy & Applied Research, Barcelona, Spain

Intimate partner violence (IPV) has profound and widespread health and economic implications at an individual, familial, and societal level. Violence risk assessment measures offer an evidence-informed approach to ascertain the degree of threat an abuser poses, transparent and defensible indicators for intervention and treatment decisions, and can be used to inform

---

ONLINE TABLE: Detailed summaries of the 39 studies reviewed in this article can be found in the table available online at http://www.springerpub.com/pa. Click on the link to "The Partner Abuse State of Knowledge Project" and go to Topic 15 in the online document.

professionals, perpetrators, and victims alike regarding the nature and intensity of services required to help prevent IPV. This article summarizes the state of knowledge regarding risk assessment for IPV through a systematic examination of all English publications from westernized nations from 1990 to 2011. Three search engines—PsychINFO, ScienceDirect, and Social Sciences Citation Index—identified 3,361 potentially relevant articles. After dropping duplicates and removing articles that did not explicitly examine risk assessment for IPV, 39 articles remained. Several themes emerged: (a) There is a relatively small body of empirical evidence evaluating risk assessment measures in the context of IPV; (b) continued advancements are needed in the methodological rigor of the research; (c) investigations should expand cross-validation research to diverse samples (e.g., gay, lesbian, bisexual, and transgender [GLBT]; male victims/female perpetrators); and (d) an exciting development in IPV risk assessment research is evidence that risk assessments can serve to reduce risk levels (Belfrage et al., 2011). In terms of clinical implications, the review demonstrated considerable promise of several measures but generally reveals modest postdictive/predictive validity. Limited evidence for the superiority of IPV specific risk assessment measures over general violence risk assessment measures was revealed; however, this may largely be a reflection of study limitations. Given the challenges in comparing across studies and the heterogeneity of partner abusers, it seems premature to recommend one preferred assessment measure/approach to clinicians.

Intimate partner violence (IPV) represents a profound public health problem because of its sheer prevalence and the associated personal, social, and economic costs to the individual, the family, and society.[1] According to a recent review of IPV research from English publications spanning the past 10 years, approximately one in four (23.1%) women and one in five men (19.3%) reported experiencing physical violence in an intimate relationship (Desmarais, Reeves, Nicholls, Telford, & Fiebert, 2012a). In a parallel review of IPV perpetration by men and women that examined 111 articles from the same decade, it was concluded that one in four women (28.3%) and one in five men (21.6%) in developed nations report perpetrating physical abuse against an intimate partner (Desmarais, Reeves, Nicholls, Telford, & Fiebert, 2012b). In fact, domestic violence is a greater problem in the criminal justice system than any other form of violence, including sexual assault, robbery, and stalking (Hart, 2009).

Spousal assaults account for as much as 12% (Department of Justice Canada, 2012) to 50% of all violent crimes reported to police in Canada (Hart, 2009). In addition to the personal and social consequences associated with IPV, the partial estimated annual cost of domestic violence in Canada is in excess of $4.2 billion (Day, 1995; Greaves, Hankivsky, & Kingston-Riechers, 1995) to 7.4 billion (see Beeby, 2012) annually.

The original Justice Canada report was not available online at the time of publication. Reeves and O'Leary-Kelly (2007) reported that IPV victimization negatively affected employee work outcomes of both male and female employees. The highest organizational cost was caused by distraction from work as a (logical) consequence of the abuse (M = \$5,088 for males and M = \$3,535 for females). The cost of domestic violence to society and to male or female partners or ex-partners has also been estimated for the year 2001 for England and Wales at £23 billion (Walby, 2004). The total cost of £23 billion was reflected by (a) costs for services (criminal justice system: £1 billion a year, health care: £1.2 billion, social services: £25 billion, housing: £16 billion, civil legal services: £3 billion), (b) economic output (i.e., the costs of time off work because of injuries: £2.7 billion), and (c) human and emotional costs (£17 billion); these costs are based on a "willingness-to-pay" approach, which places an economic value on what individuals would be willing to pay to avoid the negative event (Walby, 2004, p. 92).

Compared to other violent crimes, IPV is also remarkable because of its often repetitive occurrence. Abuse in intimate relationships is frequently recidivistic (J. C. Campbell, 1986; D. G. Dutton & Kropp, 2000; Gondolf, 1997; Quinsey, Harris, Rice, & Cormier, 1998) or processional (see Lempert, 1996); meaning that unlike other violent offenses that tend to involve single events, IPV may occur daily (Sullivan, McPartland, Armeli, Jaquier, & Tennen, 2012) and/or last for decades. For instance, Hilterman (2012) found that 35% of women in a Spanish sample (N = 46) had been with their abusive partners for ≥10 years. In addition, the severity and/or frequency of abuse may increase over the duration of an abusive intimate relationship (cf. Cavanaugh & Gelles, 2005).

THE NEED FOR VIOLENCE RISK ASSESSMENTS IN THE CONTEXT OF INTIMATE PARTNER VIOLENCE

Given the dire consequences associated with IPV, it is perhaps not surprising that the lay public, justice professionals, mental health professionals, and victim advocates alike often express disbelief and frustration when victims of IPV remain with or choose to return to their abusers. Despite the misgivings of individuals who endeavor to support women attempting to achieve safety and end IPV, there is a plethora of research to demonstrate the many reasons that a victim may choose to stay in an abusive relationship or return to an abusive partner (Nicholls, Hilterman, & Tengström, 2010). Many victims are very clear in their desire and intention to leave a partner who perpetrates IPV. Increasingly, however, the extant literature demonstrates that some victims are highly committed to achieving nonviolence with their current partner. Although the dominant cultural script has been firmly associated with the expectation that the only plausible reaction to IPV is to direct women to end abusive relationships, the research does not clearly indicate that the only method of achieving safety is to end the relationship. The notion that abuse *always* escalates has been largely unsupported. An important proportion of relationships with an abusive episode(s) do not increase in the frequency or severity of abuse (e.g., Cavanaugh & Gelles, 2005) and, in fact, many do not evidence recurrent episodes

(Feld & Straus, 1989). Such findings have important implications for negotiating appropriate safety planning and treatment strategies that consider the individual needs of the woman and that respect her preferences and the context of her circumstances. The decision to end an abusive relationship may not be straightforward. Henderson, Bartholomew, and Dutton (1997) cautioned that, "To fully understand the dynamics of abusive relationships, it will [also] be necessary to look at both partners" (p. 187). Women often are highly invested in their relationships and thus remain committed to the relationship and motivated to work toward safety and ending the abuse while remaining with the abuser (Nicholls et al., 2010). Unique to other violent crimes, victims of IPV often have ongoing emotional attachments and intricate social interconnectedness with their perpetrator and thus experience ambivalence and distress when confronted with violence in their relationships and the difficult decisions about how to best protect themselves (and any children). Moreover, evidence firmly demonstrates that the degree of risk associated is often exasperated when a woman threatens or attempts to leave an abusive partner (J. C. Campbell et al., 2003). Nevertheless, the rate of reoffense among IPV abusers (even among men who have participated in treatment) is higher than other violent criminals (D. G. Dutton & Kropp, 2000). Ultimately, regardless of whether a woman decides to stay with an abusive partner or to end an abusive relationship, mental health and justice professionals often have the responsibility of attempting to determine the level of risk posed by the abuser and put into place appropriate management and treatment options.

One effective strategy of enhancing safety and improving services delivered to victims and perpetrators of IPV is to perform violence risk assessments. IPV risk assessments can be used to determine the likelihood of abuse, or repeat abuse, and the form and severity that future abuse might take, including the risk of femicide. Risk assessments with IPV perpetrators represent a means of informing safety planning, justice interventions, sentencing, and treatment decision making.

What Is a Violence Risk Assessment?

The term *risk* refers to a hazard that is to be identified, measured, and ultimately prevented (Maden, 2007). Hart (1998) defined risk as a hazard that by definition is unknown and therefore can only be predicted with uncertainty. This reflects the fact that risk is a moving target, not static; it is dynamic, which means it changes across time and situations. By definition, an *assessment* is the process of gathering information to make decisions. The information that is accessed will depend on the context, resources available, and the nature of the assessment. Kropp, Hart, and Belfrage (2005) define *spousal assault risk assessment* as "the process of gathering information about people to make decisions regarding their risk of perpetrating intimate partner violence" (p. 2).

In general, the more sources of information the assessor is able to access, the more likely the assessment will be thorough and accurate.[2] Best case scenario, an assessor is able to include personal interview(s) with the individual of interest and collateral

informants (family, prior treatment providers, victims), psychological testing and/or medical testing (as appropriate), a review of clinical files, and often a consideration of administrative records (e.g., criminal record; Heilbrun, 2001; Risk Management Authority, 2006). Kraemer and colleagues (1997) provided a succinct and clear definition of risk assessment: "The process of using risk factors to estimate the likelihood (i.e., probability) of an outcome occurring in a population" (p. 340). Ultimately, the information is then used to prevent an adverse event from occurring or to reduce its likelihood or the potential harm. Formal violence risk assessments typically are conducted by forensic mental health professionals (e.g., psychologists or psychiatrists with additional training and expertise in forensic mental health issues); however, many direct care and service providers of diverse allied professions are involved in risk assessments (nurses, police, social workers), and preliminary research suggests they produce valid assessments of IPV risk (Storey, Gibas, Reeves, & Hart, 2011).

To offer an opinion of risk, the assessor must have a clear understanding of the behavior of interest. Experts have defined *violence* as "the actual, attempted, or threatened physical injury of another person that is deliberate and non-consensual" (Webster, Douglas, Eaves, & Hart, 1997, p. 24). Dr. Christopher Webster[3] and coauthors of one of the most widely used and well-validated risk assessment tools, the Historical-Clinical-Risk Management-20 (HCR-20) violence risk assessment scheme (Webster et al., 1997),[4] stipulate that violence is an action that is obviously likely to cause harm to another person; however, behavior that would be fear inducing to the average person may also be counted as violence. Webster et al. (1997) further caution that violence is not defined by the resulting damage experienced by the victim. Similarly, assessors should be mindful that spousal assaults typically occur in the context of strong emotional attachments, and thus IPV that results in little or no physical harm may still have dire psychological consequences. That being said, legal definitions of what constitutes an assault do not predominantly reflect psychological and emotional harm, and for that reason, IPV risk assessment measures generally do not include psychological violence in their operationalizations of IPV. For instance, the Spousal Assault Risk Assessment (SARA) Guide (Kropp, Hart, Webster, & Eaves, 1999) and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp et al., 2005) define *spousal assault* or *IPV* as the "actual, attempted, or threatened physical harm of a current or former intimate partner" (Kropp et al., 2005, p. 1).

Approaches to Violence Risk Assessment

Unstructured Clinical Judgment. Traditionally, clinicians have conducted risk assessments based entirely on their professional judgment. In many jurisdictions, unstructured clinical risk assessments are gradually being replaced by assessments that rely on empirically validated measures to provide structure, transparency, replicability, and validity to violence assessments.

This approach is referred to as "unstructured" because there are no constraints or guidelines followed by the evaluator. Decisions are based entirely on the discretion of

the assessor and justified by their training, expertise, and professional designations. The unstructured clinical approach has been summarily dismissed as an "informal, in the head, impressionistic, subjective conclusion, reached (somehow) by a human clinical judge" (Grove & Meehl, 1996, p. 294). Devastating critiques of unstructured clinical judgment (UCJ) include wide criticism in the violence risk assessment field of the very limited accuracy, vulnerability to heuristics and biases, and poor documentation associated with the approach. The major weaknesses include the limited reliability, validity, and accountability (Litwack & Schlesinger, 1999; Quinsey et al., 1998; for a discussion specific to IPV, see Nicholls, Desmarais, Douglas, & Kropp, 2007). There is now widespread denunciation of unstructured violence risk assessments, with leading authorities concluding that "unstructured clinical judgment by itself is no longer a useful or necessary approach to appraising violence risk" (Heilbrun, Yasuhara, & Shah, 2010, p. 5).

Actuarial Risk Assessment. As the field moved toward increasing the structure of risk appraisals in the hopes of improving clinical accuracy, several groups began to investigate the extent to which actuarial risk assessments might improve the plight of the field. Actuarial assessments are a "formal method" (Heilbrun et al., 2010) using fixed and explicit procedures. The actuarial approach combines risk factors through the use of explicit rules (i.e., algorithm or equation; Grove & Meehl, 1996; Meehl, 1954/1996). Actuarial measures involve the use of statistical models and previous research on risk factors and recidivism to create formulas that provide a probability that future violence will occur.[5] When using the tool, the assessor indicates if the risk factors listed are present or absent (or if relevant, scores them according to the guidelines provided), then adds up the values to get a total score. The total score corresponds to a specific level of risk for future violence (e.g., 50%) over a given period (e.g., the next five years). There is now "a 50+ year history of research comparing the accuracy of unstructured clinical judgment to actuarial approaches, with a consistent, modest advantage in predictive accuracy for the later" (Heilbrun et al., 2010, p. 5).

Although the actuarial approach appears to address the main problems of unaided clinical prediction (i.e., low reliability and validity), it suffers from other limitations. For instance, algorithms optimized in any given research sample are destined—almost axiomatically—to degrade in accuracy on cross validation. Actuarial models often fail to include low base rate but potentially vital case-related information (e.g., homicidal ideation or intent). Many evaluators are uncomfortable with the lack of practical use often associated with many actuarial assessment measures (i.e., lack of attention to case-specific risk variables and the focus on prediction vs. management and prevention; Douglas & Kropp, 2002; Hart, 1998; for discussion specific to IPV, see Nicholls et al., 2007).

The distinctive feature of "nondiscretionary" approaches (Hart & Logan, 2011) to risk assessments is that the ultimate decision reflects fixed and explicit rules (Meehl, 1954/1996). Despite the notion that these approaches are entirely objective, experts remind us that discretion is evident in all violence risk assessments to some extent

(Hart & Logan, 2011). Even when using nondiscretionary risk assessment procedures, professional judgment is necessary to determine what approach and specific measure(s) is used, and the individual must still administer, score, and interpret the measure (Hart & Logan, 2011).

Structured Professional Judgment. The limitations associated with both and pure actuarial approaches ignited the development of a model that attempts to reflect the strengths of both approaches (discretion and relevance to treatment on the clinical side and satisfactory reliability and predictive validity on the actuarial side) while at the same time minimizing the weaknesses of each approach. This model is called *structured professional judgment* (SPJ; Borum, 1996; Douglas & Kropp, 2002; Hart, 1998). Kropp and Hart (2000) defined SPJ as "a decision made without fixed and explicit rules but based at least in part on consideration of a standardized information base" (p. 102).

SPJ instruments might best be considered professional guidelines or *aides-mémoire* rather than formalized tests, although their evaluation requires formal analysis of reliability and predictive validity. SPJ is an attempt to bridge the gap between UCJ and actuarial assessments (Douglas & Kropp, 2002; Hart, 1998). In the SPJ model, risk factors are critically examined, combined, and integrated to reach a conclusion. The SPJ approach is more prescriptive and systematic than UCJ.

SPJ tools not only let clinicians integrate their own judgment but also provide a minimal structure by supplying the assessor with a list of relevant risk factors culled from consultation, theory, and the literature to consider. SPJ tools suggest what risk factors you should consider, but you do not add up those factors to get a final score. Instead, the final risk decision is left to the assessor (e.g., low, moderate, or high risk). Four studies have directly compared the actuarial and SPJ approach. Heilbrun and colleagues (2010) reviewed the literature and concluded that the extant literature suggests that actuarial and SPJ approaches are generally comparable in the degree to which they provide predictive accuracy for violent outcomes.

Who Can Conduct Violence Risk Assessments?

There is some variability in the literature regarding who can and should conduct violence risk assessments. For instance, Mills, Kroner, and Morgan (2011), Canadian experts on violence risk assessment, asserted that "violence risk assessment requires graduate level-training or equivalent knowledge" (p. 4). The Canadian Association of Threat Assessment Professionals Website (http://www.catap.org/blue-resources/FAQ.htm#tools) recommends that individuals using SPJ violence risk assessment tools have "expertise in individual assessment such as formal training and/or work-related experience in psychological, psychiatric, social work, nursing, or correctional assessment." Guiding authorities also include the American Psychology-Law Society Specialty Guidelines, American Academy of Forensic Psychology, and American Psychological Association's (APA) Ethical Principles of Psychologists and Code of

Conduct (APA, 2002, 2010). *Competence* is defined according to APA's amendments to the Ethical Guidelines as "Psychologists provide services, teach, and conduct research with populations and in areas only within the boundaries of their competence, based on their education, training, supervised experience, consultation, study, or professional experience (p. 4). Specifically, section 2.01(f) requires that when assuming forensic roles, psychologists are or become reasonably familiar with the judicial or administrative rules governing their roles" (p. 5).

To summarize, the field has often concluded that not only should a violence risk assessor be a mental health professional—not all mental health professionals are believed to have the required skills and expertise to complete a violence risk assessment. In addition, the authors of specific measures have set out specific guidelines regarding the qualifications required to use the various measures available in the field. Several measures used to inform violence risk assessments require specific professional credentials. For instance, in the HCR-20 Violence Risk Assessment Guide, Webster and colleagues (1997) note that users require a psychology or relevant degree plus relevant test administration training. However, some measure developers have attempted to define user qualifications more broadly to support a more universal implementation of these approaches into clinical practice (e.g., B-SAFER; Kropp et al., 2005). Given the cost of employing mental health professionals to conduct risk assessments and the low base rate of most of the outcomes of concern (i.e., the rarity of homicide and serious incidents of violence), many institutions and organizations extend risk assessment responsibilities to diverse mental health (e.g., nurse) and allied professionals (e.g., police, social workers). In support of including advocates and diverse direct care providers in risk assessments, Storey et al. (2011) recently concluded that criminal justice professionals (mostly police) can be trained to conduct IPV risk assessments. That being said, many of the published measures provide qualification requirements; these are articulated in the following text with respect to each measure.

Risk Assessments for Intimate Partner Violence

There are now more than 11 risk assessment measures developed specifically to assess and manage IPV offenders, in addition to a small number of general violence risk assessment tools and various approaches (e.g., victim risk assessment, pilot measures) to assessing risk of IPV that have been researched with spousal assaulters to date. A handful of prior reviews of IPV risk assessment measures have been completed, and it is therefore appropriate to comment on the findings from those reviews and meta-analyses and how this article builds on and adds to the existing body of knowledge. To our knowledge, the first review (D. G. Dutton & Kropp, 2000) of IPV risk assessment focused exclusively on two IPV specific risk assessment measures published at the time (Danger Assessment Scale [DA], J. C. Campbell, 1995; SARA, Kropp et al., 1999). The next major review was a meta-analysis of the predictive accuracy of approaches and tools used to assess the risk of recidivism in

male IPV perpetrators (Hanson, Helmus, & Bourgon, 2007). Hanson and colleagues (2007) examined 18 studies and concluded that tools designed to predict general or violent recidivism (average weighted d of .54, 4 studies) showed a degree of accuracy that was highly consistent with that of measures developed to assess for IPV specifically (average weighted d of .40, 10 studies). Of particular note, female victims' assessments of their partners' risk levels were also well within the same range (average weighted d of .36, 5 studies; Hanson et al., 2007). This study builds on the findings of Hanson et al. (2007) by extending the review by an additional three years and reporting data for an additional 24 studies. Our review is also unique because we examined only articles published in peer-reviewed journals, whereas the Hanson meta-analysis included government reports and conference presentations, for instance.

Following the Hanson et al. (2007) meta-analysis, a report published by the Alberta Institute of Health Economics (IHE; Guo & Harstall, 2008) looked at the prevention of spousal violence through the use of risk assessment measures. The IHE document was restricted to (a) studies that *sampled males*, (b) who had *contact with the police* system, and (c) that compared two or more risk assessment instruments published between 1995 and 2007 in English or German and (d) did not include women's perceptions of their partner's future risk. The IHE (2008) report identified only eight of the 18 studies (from 33 papers) included in the earlier Hanson meta-analysis, although they employed a broader search strategy by examining government Websites and Google, for instance. We also feel the inclusion of the 7 published studies that report on victim assessments is an important contribution to this body of work.

The third and most recent review of IPV risk assessment measures was published by Bowen (2011), but that publication did not include non–IPV-specific violence risk assessment measures ($k = 9$ studies covered in our review; e.g., the Violence Risk Appraisal Guide [VRAG]; Harris, Rice, & Quinsey, 1993; Quinsey, Harris, Rice, & Cormier, 2006) or studies reporting on pilot measures ($k = 3$ studies covered in our review).

The Present Review

Our review of 39 published validation studies is anticipated to be more comprehensive than some prior reviews (e.g., the IHE report evaluated eight studies). We anticipate this review will add additional information to understanding the state of the field of IPV risk assessment by expanding the range of measures and approaches considered. For instance, the Bowen review examined only IPV measures, and the IHE review discussed just six different measures. Our review includes IPV risk assessment measures, general violence risk assessment measures tested with abusers or in the context of IPV recidivism, as well as including pilot measures and victim reports while restricting our scope to only published, peer-reviewed studies. In total, we present a synthesis of 19 different measures and approaches to IPV risk assessment.

As part of the larger Partner Abuse State of the Knowledge (PASK) project (this journal), our intention was to provide a systematic review of risk assessment measures that have been tested in relation to their ability to predict IPV. The objectives were to: (a) identify all IPV risk assessment measures and related approaches (e.g., general violence risk assessment measures that have been evaluated with IPV offenders, reabuse, and recidivism; pilot tools; women's assessments of their partner's risk); (b) describe briefly the purpose, development, and use of the various measures; (c) provide readers with a summary of the objectives, methods, results, and strengths and limitations of the studies; and (d) summarize the psychometric properties of these diverse approaches (interrater reliability [IRR], concurrent validity, predictive validity) and provide recommendations for research and practice based on our synthesis of the extant empirical literature.

METHODS

Eligibility Criteria

Studies were included in this review if they met the following criteria. First, only studies from peer-reviewed journals were included. Second, all studies had to examine the risk of IPV perpetration in adults, but it did not have to be the main focus of the article. Third, the articles also had to provide empirical data about the measure and/or method of assessment more specifically information on validity or reliability, preferably. Because the body of work on IPV risk assessment is not large in comparison to many of the subjects covered in the PASK project and in the interest of providing comprehensive coverage of the topic, the decision was made to also include articles if they reported relevant reliability or concurrent/convergent validity data in the absence of postdictive/predictive validity. Most studies used specialized IPV risk assessment instruments (e.g., the SARA; Kropp, Hart, Webster, & Eaves, 1995, 1999). Papers that reported IPV results from studies of general risk assessment measures and pilot measures were included if they examined the measures' relationship with IPV and also met the other criteria for inclusion. Fourth, in keeping with the scope of the larger PASK project, we only examined studies conducted in Westernized nations. All studies had not necessarily been conducted in English-speaking countries, but the articles had to have been published in English.

Search Method

A systematic search of peer-reviewed journal articles published in English was conducted in PsycINFO, ScienceDirect, and Social Sciences Citation Index. We searched for all IPV risk assessment publications available online as of December 2011.[6] As Guo and Harstall (2008), noted the notion of IPV risk assessment and variables relevant to IPV recidivism really only emerged in the past two decades (also see

Hilton et al., 2004). Our results revealed 39 papers published between 1990 and 2011. Our search consisted of terms selected from four clusters:

- Intimate partner violence: partner violence or partner abuse or domestic violence or intimate partner violence or wife abuse or family violence or femicide or spouse abuse or physical abuse, risk assessment and validation
- Measurement: Test Validity or Statistical Validity or Test Reliability or Statistical Reliability or Accuracy or Predict*
- Risk assessment: Actuarial or Risk assessment or Structured professional judgment or Risk factors or Rating scale or Tool or Measure* or Instrument
- Risk: recidivism or risk or dangerousness

We used combinations of these four clusters to hone or broaden the scope of our search results as necessary to thoroughly review the available literature. ScienceDirect yielded more than 6,000 titles, even when all four clusters separated by the connector "AND" were combined, so the qualifier "and not (child OR youth)" was added. In addition, we searched specifically for the following instruments: Domestic Violence Risk Appraisal Guide (DVRAG), DA, Kingston Screening Instrument for Domestic Violence (KSID), Ontario Domestic Assault Risk Assessment (ODARA), SARA, B-SAFER, and Domestic Violence Screening Instrument (DVSI). The VRAG, Level of Service Inventory (LSI), and HCR-20 are broader violence risk assessment instruments but were included in our search and paired with the IPV cluster of terms. Throughout the search process, other systematic reviews and meta-analyses were used to verify that the search had not overlooked any key terms or publications (Bowen, 2011; Guo & Harstall, 2008; Hanson et al., 2007).

Review Process

The review process began with a search of the 3,361 resulting titles from each of the databases (PsycINFO = 722; ScienceDirect = 1,137; and Social Sciences Citation Index = 780). One of the researchers conducted this initial portion of the search (MP), determining that 286 articles were relevant for the abstract review, a process which was completed by all four members of the team (TN, MP, KR, and EH). These abstracts were put into a spreadsheet to allow the entire team to comment and come to a consensus on the inclusion of each publication. This process yielded 39 articles that are included in this review. The publications do not necessarily each represent primary studies; on occasion, a parent study is identified or two separate publications may report results from the same or overlapping databases. When this occurred, we included the separate publications but report the overlap in the samples.

Data Synthesis

Data was extracted from each article into a table, as set out by the PASK project, J. Hamel and J. Langhinrichsen-Rohling, including a column for the full reference, sample size and sample characteristics, methodology and study design, and results.

To summarize and synthesize the information and systematically pull relevant details from each article, further categories were added. For instance, the objective set out by the authors of each article was added to the table. Additional subheaders were included to describe the sample: age, ethnicity, gender of participants, education/employment information, and study setting. Details about the methods and design were included as was information on procedures, measure(s) used, how the measure was administered (i.e., interview or file based; contacts with collaterals), and study limitations. The results column was also further subdivided to include information on recidivism rates, total scores of risk assessment instruments, statistics on IRR, internal consistency, and predictive validity. The results were coded to address the properties of the risk assessment instrument(s) discussed in the article rather than the results of the authors' objectives. This table is available online at (http://www.springerpub.com/pa).

The 39 articles meeting our inclusion criteria examined the relationship between IPV and 19 separate methods for assessing IPV; 14 of which were IPV-specific risk assessment tools, 3 of which were general violence risk assessment measures, and 2 of which were related approaches (victim reports, clinical judgment). The DA ($k = 11$) was the risk assessment measure considered in the largest number of publications. That was followed by the SARA ($k = 7$), victim reports ($k = 7$), the ODARA ($k = 6$), the Level of Service Inventory-Ontario Revision (LSI-OR; $k = 4$), and the DVSI ($k = 4$). The Propensity for Abusiveness Scale (PAS; $k = 3$), the Psychopathy Checklist-Revised (PCL-R; $k = 3$), and the VRAG ($k = 3$) were mentioned in three separate studies. The B-SAFER ($k = 2$) and the Domestic Violence Supplementary Report (DVSR; $k = 2$) were discussed in two publications. Finally, the DVRAG ($k = 1$), the KSID ($k = 1$), the Partner Abuse Prognostic Scale (PAPS; $k = 1$), and the Domestic Violence Evaluation (DOVE; $k = 1$) were mentioned in just one study each. Our search also yielded three separate pilot instruments for IPV risk assessment that reported on validity. Articles that included reliability and validity information on the following risk assessment methods were also included: unstructured victim report methods ($k = 7$), clinical judgment ($k = 1$), and the combination of structured judgment and victim report ($k = 1$). Several publications reported psychometric properties for multiple measures and as such are discussed in relation to each approach/instrument later in this article. For instance, Hilton and Harris (2009) reported data for the DVSR as well as the ODARA; and Heckert and Gondolf (2004) supplied empirical data for the KSID, victim reports, and SARA. Thus, for that reason, a manuscript may be discussed in multiple sections of the results.

RESULTS

Unstructured Professional Judgment or Unaided Clinical Judgment

The predictive accuracy of unaided clinical risk assessments have not been the focus of research in the IPV field. In this review, we did not find any studies that examined UCJ; however, research suggests that many, if not most, professionals continue to rely on their intuition and subjective judgment despite the limitations of UCJ noted

previously and the fact that there have been many advances in violence risk assessment in the past two decades (Monahan et al., 2001; Otto & Douglas, 2010).

Actuarial Risk Assessment Measures—Intimate Partner Violence Specific Measures

The Ontario Domestic Assault Risk Assessment (Hilton et al., 2004) and the Domestic Violence Risk Appraisal Guide (Hilton, Harris, Rice, Houghton, & Eke, 2008). The ODARA was developed in collaboration with the Ontario Provincial Police (OPP; Rice, Harris, & Hilton, 2010). It is intended to be used by police officers and other frontline workers on the scene of an IPV call (Rice et al., 2010). The ODARA has 13 items and requires assessors to collect information such as the criminal background, IPV history, and general antisocial behavior of a male perpetrator to assess the risk of physical assault to a female victim (Rice et al., 2010). The authors suggest that on the scene of an IPV assault, this actuarial instrument can be scored and used to make decisions regarding arrest and laying charges (Rice et al., 2010). Rice and colleagues (2010) also recommend that the ODARA can be used to inform decisions and dispositions made by the court such as bail and conditional release.

The DVRAG (Hilton, Harris, Rice, et al., 2008) was developed to combine the information gathered on scene using the ODARA with an assessment that entails clinical information to produce an instrument better able to predict IPV recidivism (Hilton et al., 2004; Rice et al., 2010). As a result, the DVRAG is made up of the 13 items on the ODARA as well as a PCL-R (Hare, 1991, 2003) score (Hilton et al., 2004). The PCL-R is an instrument used to measure psychopathy rather than to assess risk; however, psychopathy as a risk factor has proven to have a robust association with violent recidivism (see Hare, 2003; Salekin, Rogers, & Sewell, 1996) and more specifically IPV recidivism (see Grann & Wedin, 2002; Hilton, Harris, & Rice, 2001).

Six peer-reviewed published studies employing the ODARA (Hilton et al., 2004) were identified in our literature review (see Table 1). All six of the studies were published by the authors of the measure. The development sample (Hilton et al., 2004) was based on 589 male offenders drawn from Ontario police databases plus an additional 100 cases for cross-validation. Only cases involving a male who evidenced forceful physical contact against his current or former wife or common-law wife based on victim reports or police evidence were included in the sample. This was a retrospective study with an average follow-up of 4.79 years after the index offense (SD = 1.08). Recidivism was defined as any subsequent violent assault known to police that was committed against an (ex-) wife or (ex-) common-law wife regardless of whether charges were laid. The final constructed ODARA (via setwise and stepwise selection analyses) yielded an area under the receiver operating characteristic curve (AUC)[7] of .77 (SE = 0.02, CI = ±0.04). The correlation[8] between the ODARA and dichotomous wife assault recidivism was moderate (r = .43, p < .001). Positive predictive power ranged from .30 to .72, and negative predictive power ranged from .70 to .96. In the cross-validation sample, the ODARA mean score was the same but the correlation with recidivism was

TABLE 1. Ontario Domestic Assault Risk Assessment (ODARA) & Domestic Violence Risk Appraisal Guide (DVRAG)

Study	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Hilton et al. (2004) *Canada* *ODARA*	Male offenders $N = 589 +$ 100 cases for cross validation	38 yrs	Retrospective follow-up (4.8 yrs)	Any subsequent violent assault against an (ex) wife or (ex) common-law wife	Construction Sample: 30% Cross-validation Sample: 26%	IRR: Correlated ODARA scores for pre-index & index info with ODARA scores for post-index info ICC = 0.90, Recidivism ICC = 0.91 Internal consistency: Not reported	DA: $r = .43^*$ SARA: $r = .60^*$ DVSR: $r = .53^*$ $^*p < .01$	Construction Sample: • AUC = .77 (SE = .02; CI +/− 0.04), d = 1.1 • Recidivism (r = .43) Cross-Validation Sample: • AUC = .72 (SE = .06) ns
Hilton et al. (2008) *Canada* *ODARA*	Sample 1: Male offenders with a history of IPV $N = 303$ Sample 2: Male offenders with a history of IPV $N = 346$	S1: 36 yrs S2: 35 yrs	Retrospective follow-up (5.0 yrs)	Any subsequent violent assault against an (ex) wife or (ex) common-law wife	S1 = 49% S2 = 41%	Not reported	SARA: $r = .60^*$ VRAG: $r = .53^*$ PCL-R: $r = .55^*$ DA: $r = .43^*$ DVSI: $r = .52^*$ DVRAG: $r = .87^*$ $^*p < .001$	Sample 1: • AUC = .67, 95% CI [0.61–0.73] Sample 2: • AUC = .65, 95% CI [0.59–0.71] • d = .55

TABLE 1. *(continued)*

Study	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
DVRAG						IRR: $n = 10$ $r = .92^{a}$* ICC $= .90$* *$p < .001$ $n = 16$ $r = .83^{b}$ (no p-value reported) Internal consistency: Not reported	ODARA: $r = .87$* SARA: $r = .63$* VRAG: $r = .66$* PCL-R: $r = .72$* DA: $r = .46$* DVSI: $r = .50$* *$p < .001$	Sample 1: • Not reported Sample 2: • AUC $= .70$, $p < .001$ • $d = .75$
Hilton et al. (2008) *Canada* *ODARA*	Female victims of IPV $N = 111$	36 yrs	Correlational Design (no follow-up)	No outcome data collected. Instead, severity of the assault was coded from file.	Not reported	IRR: Not reported Internal consistency: $\alpha = .65$	CTS: $r = .25$ ($p < .01$) 5 point injury Scale $r = .25$ ($p < .05$)	N/A
Hilton & Harris (2009)[c] *Canada* *ODARA*	IPV offenders $N = 391$	38 yrs	Retrospective follow-up ($M = 5.0$ yrs)	Wife assault recidivism documented in a police incident report or in an official criminal record	27%	IRR[c]: $>.90$ ($n = 24$) Internal consistency: Not reported	Not reported	Recidivism • AUC $= .67$, 95% CI $=$ [0.67, 0.73] • $d = .6.$

Study	Sample	Age	Design	Outcome	Base rate	Reliability	SE	Validity
Hilton et al. (2010) *Canada* *ODARA*	Incarcerated IPV offenders N = 150	30 yrs	Retrospective follow-up (8.0 yrs)	Wife assault recidivism documented in a police incident report or in an official criminal record	27%	IRR: (n = 10) At baseline: ICC = .95 At follow-up: ICC = .94 for time at risk ICC = .76 for recidivsm Internal consistency: Not reported	Not reported	Recidivism • AUC = .64, 95% CI [0.54 to 0.73] Cormier-Lang Score • r = .34, p < .05 Seriousness of criminal justice outcome • r = .24, p < .01
Eke et al. (2011) *Canada* *ODARA*	Subsample of males who murdered a female partner N = 30	37 yrs	Postdictive	Death or severe injury to victim of index offence	Not reported	IRR: r = .98 (n = 15) Internal consistency: Not reported	Not reported	Not reported

Note. IRR = inter-rater reliability; ICC = intraclass coefficient; AUC = area under the curve; CI = confidence interval; *SE* = standard error; IPV = Intimate partner violence; CTS = Conflict Tactics Scale; DA = Danger Assessment Scale; DVRAG = Domestic Violence Risk Appraisal Guide; DVSI = Domestic Violence Screening Instrument; DVSR = Domestic Violence Supplementary Report; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist – Revised; SARA = Spousal Assault Risk Assessment; VRAG = Violence Risk Appraisal Guide.

[a]This IRR calculation was based on two independent ratings of the DVRAG both blind to recidivism.
[b]This IRR calculation was based on a research assistant recoding files originally coded by a forensic clinician that did not have DVRAG training.
[c]We were unable to determine which statistic was used to calculate IRR.

smaller (base rate of recidivism 26%). The ODARA was significantly positively corre-lated with the sum of victim injury scores for subsequent IPV offenses, the sum of the Cormier-Lang Scale scores for subsequent IPV offenses, and the number of subsequent IPV incidents with acts of severe violence (Hilton et al., 2004).

The predictive validity of the ODARA in male offenders ranged from an AUC of .64 to .77 (recidivism ranged from 26% to 49%). When compared to other IPV risk assessment measures, the ODARA was significantly correlated with the SARA, DA, and DVSR (Hilton et al., 2004). There was no consistent finding regarding whether the ODARA performed better than the other measures. In Hilton, Harris, Rice, et al.'s (2008) study, the ODARA and psychopathy, as measured by the PCL-R, were the only significant predictors of recidivism. In the original cross-validation study, only the ODARA signifi-cantly predicted recidivism, although the SARA, DA, and DVSR significantly predicted recidivism in the development sample (Hilton et al., 2004). The ODARA performed better than the LSI-OR in predicting DV recidivism, the LSI-OR performed better in predicting violence of unknown victim–offender relationship, and both predicted the occurrence of any postrelease charges equally (Hilton, Harris, Popham, & Lang, 2010). The internal consistency of the ODARA was reported by Hilton, Harris, and Holder (2008). Cronbach's α was .65 based on reports of 111 women.

Two studies used the ODARA in unique populations: cases of femicide or attempted femicide and female victims of IPV (Eke, Hilton, Harris, Rice, & Houghton, 2011; Hilton, Harris, & Holder, 2008). In a retrospective study Eke and colleagues (2011) explored the ability of the ODARA to identify potential femicide perpetrators in a group of males who had attempted or committed femicide. The ODARA was coded for 30 of 146 cases of males who committed femicide or attempted femicide because these were the ones with enough detailed information available. Of these 30 cases, 43% had an identifiable index offense (i.e., the assault directly preceding the femicide or attempted femicide), which may have inflated the ability of the ODARA to identify potentially lethal cases. All but one of the male offenders fell into the highest bin on the ODARA.

The second study using the ODARA in a unique population explored its use to profile women's level of risk and severity of injuries in women attending partner assault clin-ics (Hilton, Harris, & Holder, 2008). Participants were 111 women (71% larger com-munity, 29% smaller community) who were victims of domestic violence. Women were assessed using the ODARA during the routine face-to-face nursing assessment. The outcome was not recidivism but the severity of injuries (5-point scale) and the women's own assessment of risk for future violence. The correlation between the ODARA and the Revised Conflict Tactics Scale (CTS2) (Straus, Hamby, Boney-McCoy, & Sugarman, 1996) was $r = .25$, between the ODARA and sexual assault was $r = .22$, between the ODARA and the 5-point injury scale was $r = .25$, between the ODARA and the pres-ence of potentially lethal acts was $r = .19$, and between the ODARA and prior medical treatment for assault by the perpetrator was $r = .26$ (all significant at $p < .05$). Victims' perceptions of future violence were not significantly related to the ODARA. Participant ratings were most strongly associated with reports of increasing severity of violence ($r = .29, p < .05$) but not with perceived increase in frequency of assault.

IRR was based on correlating ODARA scores for pre-index and index information with ODARA scores for post-index information for 24 cases coded by two research assistants (double blind; Hilton et al., 2004). The resulting intraclass correlation coefficient (ICC)[9] for the pre- and post-index scores was .90, and the ICC for the ODARA score and recidivism was .91 (Hilton et al., 2004). In a field simulation, two police officers who were not involved in the ODARA construction independently scored the ODARA for 10 cases (ICC of .95, $p < .001$; Hilton et al., 2004).

Lastly, the authors of the ODARA also created the DVRAG to include more clinically relevant information in the prediction of IPV (Hilton, Harris, Rice et al., 2008). Hilton and colleagues (2008) did this by using corrections files to code the ODARA, VRAG, DA, SARA, DVSI, and PCL-R and then running correlations see which instrument added to the predictive validity of the ODARA. One sample used in the development of the DVRAG was made up of participants ($N = 303$) from the original ODARA development sample that had the most detailed correction records. The other sample used was made up of new IPV offenders ($N = 346$) drawn from police records in a similar manner to the ODARA construction sample. The definitions for the index offense and recidivism are the same as in the ODARA construction sample (Hilton et al., 2004). The follow-up was on average 5.01 years ($SD = 1.44$). In the first sample, the AUC for the ODARA was .67, 95% CI [0.61–0.73]. Only the PCL-R added significantly to the prediction of the number of recidivism events, the number of severe violence incidents, and for total recidivism injury. Thus, the DVRAG was a combination of the ODARA and the PCL-R total score with new weights calculated. The IRR of the DVRAG was established by two independent blind raters who coded 10 randomly selected cases. The ICC was .90 and $r = .92$. The authors also compared the scoring of a forensic clinician to a research assistant for 16 cases and the correlation was .83. Using the second sample, the AUC for the ODARA was .65, and for the DVRAG, the AUC was .70. The DVRAG represented an improvement in predictive validity over the ODARA score. When the two samples were combined, all measures significantly predicted recidivism.

In summary, the ODARA has been investigated in samples of male offenders with a known history of IPV, female victims of IPV, and male offenders who have committed femicide or attempted femicide. The ODARA predicts recidivism, it is significantly related to other measures of IPV, and has been found to appropriately classify perpetrators into various risk categories. Most of the studies reporting predictive validity are retrospective follow-up designs and thus are limited. In addition, most studies used file reviews in the absence of interviews. In addition, recidivism was determined only through official records of subsequent IPV. Those studies that did include interviews to collect information for coding the ODARA did not include recidivism data; and thus, it is very possible that incidents of IPV are drastically underestimated. Studies to date have demonstrated moderate-to-high predictive accuracy, and the measure has been found to have good IRR (see Table 1).

The Domestic Violence Supplementary Report (Ontario Ministry of the Solicitor General, 2000). The DVSR was developed by the OPP to be used by frontline police

officers to assess risk for IPV (Ontario Ministry of the Solicitor General, 2000). It is a rationally derived tool based on a literature review and consultation with experts in the field of IPV. The 19 items cover relationship problems, stalking, threats, weapons, mental illness, substance abuse, violation of supervision orders, violence toward pets, and the victim's fear. The items are coded as present or absent and then summed for a total score. The risk factors on the DVSR were chosen both for their relationship to IPV and because all of the items can be coded based on information readily available to frontline police officers. It should be completed based on a victim interview, police investigation, and the perpetrator's criminal history (see http://www.fact.on.ca/Info/dom/police00a.pdf).

There are only two published studies that report validity information for the DVSR, and these studies demonstrated fair to good predictive validity for IPV recidivism (AUCs ranging from .57 to .67, although the .57 was not significant). In the first study, Hilton et al. (2004) were developing and validating the ODARA, which was derived from the risk factors on the DVSR (for details on this study, see "The Ontario Domestic Assault Risk Assessment and the Domestic Violence Risk Appraisal Guide" section of this article). The correlation of the DVSR to recidivism was .26 ($p < .001$), and the predictive validity of the DVSR yielded an AUC of .67 (95% CI [0.63–0.71]). The correlation between the ODARA and the DVSR was $r = .53$ ($p < .01$).

In the second study, Hilton and Harris (2009) were investigating the impact of ambiguous recidivists on the predictive ability of ODARA and the DVSR. Participants were 391 male IPV offenders whose files were obtained from Ontario police databases. The two measures were coded from file information alone, and recidivism was subsequent wife/common-law assault. The AUC for the DVSR was .59 (95% CI [0.52–0.65]) but increased when the ambiguous recidivists were discarded (AUC = .61, 95% CI [0.54–0.68]). The authors created smaller groups with a base rate of recidivism of 50%. The average AUC was .61 and the range was .57 (95% CI [0.46–0.69]) to .65 (95% CI [0.54–0.76]).

The DVSR was created for frontline police officers to use when called to domestic incidents. The most significant limitation of both of these studies is that they do not explore the validity of the tool under these specific conditions.

Propensity for Abusiveness Scale (D. G. Dutton, 1995). The PAS was constructed as a means to assess abusiveness in IPV perpetrators without eliciting reactive behaviors or behaviors aimed at social desirability that tend to minimize the perpetrators' description of the abusive situation (D. G. Dutton, 1995). Rather than directly asking about abusive behaviors, the PAS is made up of subscales that measure personality traits, anger, trauma symptoms, and family of origin issues that have been found to have robust correlations with abusiveness (see D. G. Dutton, 1995 for information on subscales). D. G. Dutton, Landolt, Starzomski, and Bodnarchuk (2001) suggested that because of the use of personality and temperament subscales in the PAS, it would be a useful tool for programs aimed at secondary prevention programs for potential IPV perpetrators. The PAS is made up of 29 items answered

using Likert scale ratings. It is to be completed by the perpetrator of IPV (D. G. Dutton, 1995). In our review of the literature, three published studies were identified that used the PAS (D. G. Dutton, 1995); two studies assessed the likelihood that a particular group of men had the propensity to commit IPV (D. G. Dutton, 1995; D. G. Dutton et al., 2001); and a third paper reported results of the test–retest reliability of the PAS (Clift, Thomas, & Dutton, 2005).

The first study (D. G. Dutton, 1995) reflected the development of the PAS and was conducted with men in treatment for IPV and their female partners. The strength of the study is that the authors included female partner reports of emotional abusiveness and compared the scores from the known treatment group to a demographically matched control group. This study did not, however, compare the scores of the PAS to recidivism (i.e., an examination of predictive validity) but rather used discriminant function analysis to classify participants based on their PAS scores on the two scales on the Psychological Maltreatment of Women Inventory (PMWI; Tolman, 1989). In addition, the study reported convergent and criterion validity of the PAS. After the initial development phase, the author then cross-validated the PAS with a second group of men in treatment and their female partners. The results of the study indicated high internal consistency ($\alpha = .92$) as well as good convergent and divergent validity. Specifically, the PAS correlated ($r = .51, p < .001$) with the Dominance/Isolation Scale and the Emotional Abuse Scale of the PMWI ($r = .47, p < .01$). It generated a significant discriminant function that correctly classified 80.0% of men on the dominance/isolation factor and 84.4% on the emotional abuse factor. In the cross-validation sample, the PAS maintained good internal consistency ($\alpha = .88$), and it correctly identified 82.2% of men on dominance/isolation and 81.3% of men on emotional abuse.

The purpose of the second study (D. G. Dutton et al., 2001) was to validate the PAS in diverse male samples. The total sample ($N = 363$) drew from four groups: a nonviolent clinical outpatient population ($n = 50$), male college students ($n = 149$), a sample of gay men in long-term relationships ($n = 104$), and a new sample of known assaultive men ($n = 60$). Inclusion criteria required all participants to be in a relationship of at least six-months duration. The men were solicited from mental health clinics, newspaper advertisements, college classes, and from treatment groups for IPV. Similar to the development and cross-validation study, there was no recidivism data published, but instead, the total score was correlated with the PMWI (Tolman, 1989). The partners of participants completed the criterion measures (PMWI and the Marshall Scale [Severity of Violence Against Women; Marshall, 1992]). The authors do not provide details of administration of the measures; and thus, it is unclear how many of the measures were completed by the male participants and partner participants. The results (D. G. Dutton et al., 2001) indicated the PAS scores of the men in the gay sample, the college sample, and the assaultive sample were significantly correlated with the Dominance/Isolation and Emotional Abuse Scales of the PMWI; this was not the case for the clinical sample. It should be noted that Figure 3 of the article indicates that the clinical sample PAS scores were significantly correlated with the scales of the PMWI, but this is not the case in Table 2. Only the

TABLE 2. Domestic Violence Screening Inventory (DVSI) & Domestic Violence Screening Inventory- Revised (DVSI-R)

Study	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/Convergent Validity	Predictive/Postdictive Validity
Williams & Houghton (2004) USA DVSI	Male offenders arrested for IPV against female partners $N = 1,465$	32 yrs	Prospective validation study	Official re-arrest records over 18 months	29% - arrested for violation of restraining order and domestic violence reoffending 53% - arrested for other types of criminal offending	IRR: Not reported Internal consistency: $\alpha = .71$	SARA total score: $r = .54$ SARA summary risk: $r = .57$	Rearrest data- DVSI total score: • Re-arrest for IPV: AUC $= .61, p = .000$ • Frequency of IPV rearrest: $r = .18, p = .00$ • Total rearrest: AUC $= .65, p = .000$ • Frequency of total re-arrest: $r = .21, p < .00$
Williams & Grant (2006) USA DVSI-R	Completed risk assessments for Individuals arrested for IPV charges $N = 14,970$ (71% male)	33 yrs	Postdictive field study	Re-assessment by Family Services as a consequence of re-arrest	25%	Not reported	Not reported	DVSI-R AUC = .71 ($p < .05$) Summary Risk Ratings: • Imminent risk to victim: AUC = .64, 95% CI [0.70–0.72] • Imminent risk to other: AUC = .61, 95% CI [0.63–0.65]

Study	Sample	Age	Design	Outcome	Base rate	IRR	Predictive validity	Results
Hilton et al. (2008) Canada DVSI-R	Sample 1: Male offenders with a history of IPV N = 303; Sample 2: Male offenders with a history of IPV N = 346	S1: 36 yrs S2: 35 yrs	Retrospective follow-up (5.0 yrs)	Any subsequent violent assault against an (ex) wife or (ex) common-law wife	S1 = 49% S2 = 41%	Not reported	ODARA: $r = .52$* VRAG: $r = .31$* PCL-R: $r = .34$* DA: $r = .36$* DVRAG: $r = .50$* SARA: $r = .53$* *$p < .001$	Sample 1: • $r = .17, p < .01$ Combined Sample: • AUC = .61, $p < .01$
Williams (2012) USA DVSI-R	IPV perpetrators N = 3,569 (70.5% Male)	35 yrs	Cross validation field study	Re-arrest within 18 months - 3 outcomes: • New family violence offences only • Violations of protective orders • New family violence offences & violation of court order	New family violence offences only: 33.8% Violations of protective orders: 5.5% New family violence offences & violation of court order: 19.1%	Not reported	Not reported	DVSI-R total score: • New family violence offences only: AUC = .62, 95 CI [0.60, 0.64] • Violation of protective order: AUC = .72, 95% CI [0.66, 0.77] • New family violence & violation of court order: AUC = .73, 95% CI [0.70, 0.75] No differences between male & female perpetrators with the exception of 3rd outcome

Note. IRR = inter-rater reliability; AUC = area under the curve; CI = confidence interval; IPV = Intimate partner violence; DA = Danger Assessment Scale; DVRAG = Domestic Violence Risk Appraisal Guide; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist – Revised; SARA = Spousal Assault Risk Assessment; VRAG = Violence Risk Appraisal Guide.

college sample PAS scores were significantly correlated with the CTS (Straus, 1979), and the assaultive sample PAS scores were also correlated with the Marshall Scales. The PAS was highly predictive of emotional abusiveness in all samples, and it significantly predicts emotional abuse, use of dominance isolation, physical abuse, and the use of threats to kill or injure.

Finally, in a third study, the test–retest reliability of the PAS was demonstrated by Clift et al. (2005). In this study, 27 male and 37 female university students completed the PAS at the initial study phase and 2 years later. The overall reliability for the PAS for women was .85 and for men was .63. In the full sample, the reliability was .77. Each of the scales demonstrated significant correlations between Time 1 and Time 2; the subscale with the highest reliability was Recalled Negative Parental Treatment, and the least reliable subscale was Trauma Symptoms, although the test–retest reliability correlation was still significant. Given that the authors sampled university students, it is unclear if the test–retest reliability will generalize to offenders and community samples.

In summary, the PAS has demonstrated a strong relationship with partner reports of emotional abuse and physical abuse in some samples and in high test–retest reliability in university students. The author asserts it has the advantage of being a nonreactive measure that allows accurate self-reporting by males of traits and behaviors predictive of abuse. Although the PAS has not been studied extensively, it has been tested in various populations, including males with a history of IPV perpetration, a sample of gay males, and even female college students. A limitation of this body of measure is that the PAS has not been examined in relation to recidivism.

Domestic Violence Screening Instrument (Williams & Houghton, 2004). The DVSI was created and first employed by the Colorado Department of Probation Services for use by probation officers (see Williams & Houghton, 2004). The initial version of this actuarial instrument consisted of 12 items, and the numerical total score ranged from 0 to 30. The questions are designed to collect information pertinent to "an offender's supervision level, including: (1) criminal history; (2) past DV, alcohol, or substance abuse treatment; (3) past DV restraining/protective orders, including violations; (3) [*sic*] previous non-compliance with community supervision; and (4) various other static and dynamic factors" (Hisashima, 2008, p. 1). After the 12 items are scored, the probation officer sums the scores and the higher the total score, "the higher the risk for reoffending, noncompliance with court, and probation orders, and thus, the higher the risk to victims" (Williams & Houghton, 2004, p. 441). A modified 11-item version of the DVSI, the DVSI-Revised (DVSI-R), was implemented statewide in Family Services and Connecticut Courts in Connecticut, United States (Williams & Grant, 2006). Readers should note that this is intended as a screening measure; therefore, a high score is interpreted as an indication of the need for a more thorough IPV assessment. For instance, in some jurisdictions, the DVSI policy indicates that high scores on this measure indicate the need for a SARA assessment (Hisashima, 2008).

In our literature search, we found three studies that employed the DVSI or its revised version, the DVSI-R. The DVSI was piloted in a prospective field study in four of the 22 judicial districts of Colorado, United States (Williams & Houghton, 2004). In the period from July 1997 to March 1998, the probation officers rated 1,465 male offenders on the DVSI. A strength of this pilot study was the inclusion of a small group of the victims ($N = 125$) to obtain the victim's information on reoffending of the perpetrator.

In this first study, the DVSI was found to have acceptable internal consistency ($\alpha = .71$). A limitation of this study was the absence of IRR. According to Williams and Houghton (2004), this was caused by the logistics of implementing the DVSI in the field. To obtain concurrent validity, a subsample of 434 cases were also rated on the SARA (Kropp & Hart, 2000). The total score of the DVSI correlated strongly with the total score of the SARA ($r = .54$). The SPJ rating of the SARA also correlated ($r = .57$) with the total score of the DVSI. Official records were used as the outcome to measure predictive validity. Two types of outcome were used in this study. The first was a combination of arrests for violations of IPV restraining orders and arrests for IPV reoffending. These two types of arrests were combined into a single partner violence reoffending outcome measure. In the 18-month follow-up period, 29% of the sample engaged in such behavior. The second outcome was a general reoffense measure composed of any reoffending during the 18-month follow-up period. The total reoffending prevalence was 53%. Both outcomes were dichotomized. The AUC was good and the correlation coefficient for the IPV outcome was significant but small (AUC $= .61, r = .18, p = .00$), both of which were similar to the general reoffending outcome (AUC $= .65, r = .21, p < .00$).

The interviews with the 125 victims were used to obtain data on victim reports of reoffending after a follow-up period of six months (Williams & Houghton, 2004). The interviews with victims were conducted by telephone using a questionnaire (a modified version of the CTS2; Straus, 1979, 1990) on three indices: index of control, index of threats, and index of severe threats. Similarly, two indices of physically violent behaviors were used: index of violence and index of very severe violence. Of the interviewed women, 35% reported that their partner used some type of physical force during the six-month follow-up period. The AUC for the prediction of the controlling behavior was not significant (AUC $= .58, p = .14, r = .13$). This was also the case for the index regarding the less threatening behavior (AUC $= .56, p = .26, r = .09$) and the least serious violent behavior (AUC $= .49, p = .92, r = .09$). For the serious threats and violent behavior, the prediction was somewhat higher and significant (AUC $= .68, p = .001; r = .22, p < .05$) and (AUC $= .65, p = .041; r = .18, p < .05$), respectively. The predictive validity for the specific domestic violence outcomes was low. This pilot study had several strong features including the prospective design, a very large sample, and additional information on reoffending from victim self-reports.

In a prospective field study by Williams and Grant (2006), the instrument was scored between September 1, 2004 and May 2, 2005 for 14,970 cases (71% male, 29%

female). A weak element in this prospective study was the operationalization of the outcome measure. The outcome measure reflected multiple assessments (25% of the 14,970 assessments) as a consequence of rearrest within the research period. IRR and internal consistency are not mentioned in the article. The DVSI-R predicted moderately the return of offenders to the Family Services (AUC = .71, $p < .05$). The summary risk ratings (SRR; one of the modifications was to add this structured clinical assessment to the instrument to lower the resistance from Family Relations Counsellors (FRCs) against the actuarial character of the instrument), regarding the imminent risk to the victim, had a significant but low predictive accuracy (AUC = .64, $p < .05$).

The most recent validity study with the DVSI-R (Williams, 2012) was also conducted at the Family Services in Connecticut and scored by FRCs. The design was a prospective field study, and the FRCs rated the DVSI-R before court sessions in the 24-hour period between arrest and initial court appearance. The sample consisted of 3,569 family violence perpetrators older than the age of 16 years. The sample included: (a) spouses or former spouses; (b) parents and their children; (c) persons who are 18 years of age or older related by blood or by marriage; (d) persons who are 16 years of age or older other than those persons in subparagraph (c) who are presently residing together or who have resided together; (e) persons who have a child in common regardless of whether they are or have been married or have lived together at any time; and (f) persons in, or have recently been in, a dating relationship. This means that the study was not specific on the prediction of IPV but used a more general definition of domestic violence. The outcome measure was rearrest within the 18-month follow-up for: (a) new family violence offenses only (prevalence = 33.80%; $N = 872$), (b) violations of protective/retraining orders (prevalence = 5.53%; $N = 100$), (c) new family violence offenses, (d) violations of court orders (prevalence = 19.09; $N = 403$), (e) outcomes (a) to (c) combined (prevalence = 44.60%; $N = 1.38$), and (f) all rearrests combined (prevalence = 52.14%; $N = 1.86$). The internal consistency of the DVSI-R was good ($\alpha = .75$). The absence of IRR is a limitation in the development of the DVSI-R. Williams (2012, citing Edens & Vincent, 2008, p. 194) reported that IRR was a part of the training process and that field reliability can be lower compared to reliability in a research or training setting, which was why it was not included.

The predictive accuracy of the DVSI-R total score for new family violence offenses only was AUC = .62 (95% CI [0.60–0.64]), regarding violations of protective/restraining orders AUC = .72 (95% CI [0.66–0.77]), and new family violence offenses and violations of court orders AUC = .73 (95% CI [0.70–0.75]). When all three outcome measures were taken together, the AUC decreased (AUC = .66, 95% [CI 0.64–0.68]). This was also the case for all combined rearrests (AUC = .66, 95% CI [0.66–0.68]).

The research to date suggests the DVSI-R has good internal consistency but there is a lack of information on reliability between raters. The predictive accuracy has been found to be low for "nonsevere" reoffending behaviors but moderate for more serious threatening and physical violence (see Table 2).

Domestic Violence Evaluation (Ellis & Stuckless, 2006a, 2006b). Research has found the risk of IPV can spike when a woman threatens to leave or end a relationship and/or that risk can be especially high immediately following separation. It is not surprising, therefore, that the rates of IPV are particularly high among divorcing couples. For instance, Beck, Walsh, Mechanic, Figueredo, and Chen (2011) reported that 59% of cases that went through divorce mediation had incidents of IPV and that, in some of these cases, the violence did not end when the divorce was finalized. In response to this high-risk time frame and the growing trend to use mediation when divorcing, the DOVE, a 19-item instrument used to assess and manage the risk of IPV during divorce mediation proceedings was developed (Ellis & Stuckless, 2006a). The DOVE is described as a risk assessment and risk management instrument for which indicators were developed in collaboration with women in battered women shelters. The measure collects information on statistically significant predictors of IPV such as past abuse and violence, substance use and the relationship, and is given privately to each partner participating in the mediation process (Ellis & Stuckless, 2006a). It is our understanding that after completion, the mediator scores each of the DOVE items and gives a risk assessment rating of low, moderately high, high, or very high based on the numerical score (Ellis & Stuckless, 2006a). Depending on the risk assessment rating assigned to the case, mediators will offer risk management tactics to assist in safety planning to curtail future violence (Ellis & Stuckless, 2006a).

Our review of the literature identified one published empirical study of the DOVE (Ellis & Stuckless, 2006b). Ellis and Stuckless (2006b) sampled a random, sequential group of 147 male ($n = 67$) and female ($n = 80$) Canadian partners who were referred to divorce mediation by a family court judge or family lawyers or who self-referred. Neither IRR nor internal consistency were evaluated in the study. Predictive validity was not available for the total score of the instrument, but it was reported at the item level.

Abuse was reported by the female partner before separation and 4 months after the couple separated. The rates of violence reduced dramatically from the preseparation to postseparation interviews as follows: physical assault (50.1%–9.7%), serious physical injury (17.2%–14.3%), emotional abuse (84.4%–69.9%), and serious emotional harm (71.3%–47.8%; Ellis & Stuckless, 2006b). The only form of violence that remained the same was sexual assault, with 9.8% of female partners reporting this before and after separation. The authors concluded that the following factors predicted male violence preseparation and postseparation: controlling behavior, mental health problems, relationship factors, substance use, anger, and conflict (Ellis & Stuckless, 2006b).

The DOVE was administered to couples by court-based family mediators prior to participation in divorce mediation limiting the generalizability. The authors do make note that in other samples, separation has been significantly associated with lethal violence against female partners (e.g., J. C. Campbell et al., 2003); although in this sample, violence reduced postseparation (Ellis & Stuckless, 2006b). Another

limitation of the Ellis and Stuckless (2006b) study is that the definitions of abuse were not specified.

Kingston Screening Instrument for Domestic Violence (Gelles & Tolman, 1998). The KSID had its impetus in the pioneering work of Dr. Richard Gelles (Roehl, O'Sullivan, Webster, & Campbell, 2005). This was intended as an actuarial measure developed to assess future IPV (Heckert & Gondolf, 2004; Roehl et al., 2005). The KSID is an unpublished risk assessment measure making it difficult to find information regarding proper implementation. It is primarily used in Connecticut to sentence domestic violence offenders (Heckert & Gondolf, 2004). The instrument comprises[10] items that address risk factors pertaining to the perpetrator of IPV such as sociodemographic information (e.g., age, marital status, employment status, income level), abuse history (i.e., witnessed abuse, previous IPV arrest, past child abuse allegations), and substance use history (Heckert & Gondolf, 2004). The measure also includes a poverty chart to assist in assessing the item on income, a formula to determine the presence of binge drinking, and a chart pertaining to severity of injury (Roehl et al., 2005).

Although there appears to be a handful of unpublished studies (Roehl et al., 2005), we found just one published article in our search that reported psychometric information on a "simulated" version of the KSID. The authors attempted to approximate the use of the measure but did not collect data on all of the items (Heckert & Gondolf, 2004). Male batterers ($N = 840$) from four U.S. cities were recruited from a database of men admitted to batterer treatment groups. Eighty-two percent of the men had the treatment group mandated by the court, whereas 18% were attending of their own volition. The study began upon intake to the batterer program. The researchers attempted to interview the men, the initial victims, and a new female partner every 3 months for 15 months. Although partners/victims for 82% of the men were contacted, only 67% of them were available for the full 15 months. The authors assert that they were able to simulate the KSID from questionnaires answered by the male IPV perpetrators and their partners while in treatment. The questionnaires contained 7 of the 10 KSID items and the information from the other 3 items which related to poverty, substance use, and past child abuse were collected in questions similar to those found on the KSID. Heckert and Gondolf (2004) noted that they were only completely missing information on 1 item (previous violation of a protection order). Reabuse was measured using the CTS and women's self-reports.

In this study, the total KSID score was not a good predictor of whether or not a perpetrator would reabuse over 15 months (AUC = .57, sensitivity = 11%; Heckert & Gondolf, 2004). Reabuse was defined as any physical assault, verbally abusive, threatening, or controlling behavior. Note that the authors do not indicate what proportion of these female victims were new partners versus the initial victims. Throughout the 15-month follow-up period, 19% of men were not abusive, 23% engaged in more than one incident of physical reassault, and 12% of men had one occurrence of reassault. The rate of false positives was 5%, whereas 27% of perpetrators were correctly classified.

Although the Heckert and Gondolf (2004) study produced little evidence to support the validity of the KSID, it is important to remember that the KSID was not administered as it was designed but rather simulated based on previously collected questionnaires; and we have just one study from which to draw results. Considerably more research on the measure is needed before firm conclusions can be drawn regarding its use.

Partner Abuse Prognostic Scale (Murphy, Morrel, Elliot, & Neavins, 2003). The initiative to develop the PAPS (Murphy et al., 2003) was born out of dissatisfaction with the existing risk prediction instruments available at the time that, which according to the authors, had important limitations (see also D. G. Dutton & Kropp, 2000). Murphy and colleagues (2003) pointed to several inadequacies in the development of other tools, including (a) insufficient empirical data on predictive validity; (b) measures that do not gather relevant information on partner abuse because they were designed to predict general violence; (c) measures designed to only assess lethal violence (J. C. Campbell, 1995); or (d) measures devised to assess an individual's potential to commit IPV, validated in samples of nonabusers (D. G. Dutton, 1995).

The PAPS is an actuarial risk assessment scale that consists of three subscales: relationship violence problem severity (7 indicators), substance use (4 indicators), aggression history (4 indicators), and two "other" indicators—unemployment and living together at time of intake. The total score is achieved by adding up the 17 indicators: 15 dichotomous and two trichotomous items. A strength of the development of the PAPS is that the authors gathered information from both perpetrators and victims. For example, information on the severity of the abuse was obtained from both aggressors and victims using the CTS2 (Straus et al., 1996). Six of the seven indicators of the subscale relationship violence problem severity were derived from the CTS2 (Straus et al., 1996). The last indicator of this subscale, the receipt for a legal protection order in the year before treatment, was obtained from additional structured interview questions.

The indicators for the second subscale include (a) the cutoff of the Alcohol Use Disorders Identification Test (AUDIT; Babor, de la Fuente, Saunders, & Grant, 1992); (b) any self-report of drug use in the 6 months prior to treatment; (c) a positive answer from the partner on the question, "Do you feel that your partner has an alcohol problem at the present time?" (Murphy et al., 2003, p. 1092); and (d) a positive response from the partner on the question, "Do you feel that your partner has a problem with drugs other than alcohol at the present time?" The four indicators of the aggression history subscale are positive answers on the following questions: (a) "Do you feel that you were abused or neglected as a child?"; (b) "Did you ever see, hear, or know about your parents having physical fights?"; (c) client self-report of any prior arrest for violent crime not involving an intimate partner; and (d) positive collateral response to the question, "Is your partner the type of person who gets in a lot of fights with people outside of the family?" (Murphy et al., 2003, p. 1092). The remaining two other indicators are unemployment (self-report perpetrator) and currently living together with the partner (partner collateral information).

In our literature search, we found one study with research data on the PAPS in which data was gathered from 95 males attending treatment for IPV perpetration and their female partners, when possible. Murphy et al. (2003) made no mention of IRR or internal consistency; however, a strength of the study was the use of multiple outcome measures:

1. Self-report: Program completers provided postintervention outcome data on the CTS2 (physical violence, severe violence, and injuries) during the 15th session (of a 16-session program) and after a 6-month follow-up.
2. Victim's report: Information on outcome from the identified victim and/or most recent victim was obtained through the CTS2 by telephone interview at the end of the program and at 6-month follow-up.
3. Criminal recidivism data was assessed through a review of criminal histories available at the electronic database for the state of Maryland.

Outcome data from one or both partners at posttreatment was available for 76 participants (80%); 22 (29%) were identified as recidivists for any physical assault, 14 (18%) for injuries, and 12 (16%) for severe violence. For 58 cases, self-report and/ or victim report was available at 6-month follow-up; 8 (14%) were coded as recidivists for any physical assault, 3 (5%) for injuries, and 3 (5%) for severe violence. Criminal recidivism through the database of the state of Maryland was obtained for 71% of the 95 participants; 12 (18%) were determined to be recidivists for the following range of charges: assault, battery, violation of a protection order, malicious destruction of property, child abuse, telephone abuse, and assault with a deadly weapon.

The subscale relationship violence problem severity correlated significantly with any physical aggression ($r = .26, p < .05$) and any severe violence ($r = .27, p < .05$) but did not obtain a significant association with any injuries ($r = .19, p > .05$) at the posttreatment recidivism evaluation. The relationship violence problem severity scale also correlated with any physical aggression ($r = .37, p < .01$) at 6-month follow-up but did not predict criminal recidivism ($r = .08, p > .05$) at the 2- to 3-year follow-up. The substance use subscale correlated with any physical aggression ($r = .29, p < .01$) and injuries ($r = .26, p < .05$) at posttreatment but was not significantly associated with any severe violence ($r = .17, p > .05$). Substance abuse was negatively associated with any physical aggression ($r = -.20, p > .05$) at the 6-month follow-up and positively with criminal recidivism ($r = .31, p < .01$) at 2–3 years.

Contrary to the authors' expectations, the subscale aggression history did not predict any of the IPV or general recidivism outcomes. The subsequent correlations with any physical aggression ($r = .17, p > .05$), any severe violence ($r = .08, p > .05$) and injuries ($r = .14, p > .05$) at posttreatment and any physical aggression ($r = -.10, p > .05$) at the 6-month follow-up, and criminal recidivism ($r = .20, p > .05$) at 2–3 years were not significant.

The Total Prognostic Indicators Scale and the PAPS total score significantly predicted all five outcomes. The total score was moderately associated with any physical aggression ($r = .41, p < .01$) and somewhat less with any severe violence

($r = .35, p < .01$) and injuries ($r = .31, p < .01$) at posttreatment. Finally, the PAPS total score had a small but significant correlation with any physical aggression ($r = .24, p < .05$) at 6-month follow-up and criminal recidivism ($r = .23, p < .05$) at two to three years.

The PAPS is an actuarial risk assessment instrument mainly composed of static indicators and consequently contains very little information to guide interventions and risk reduction strategies. The selection of the items appears to have been partially based on the assumption that past violence predicts future violence. Also, the inclusion of the aggressor's alcohol problem before intake provided mixed results in the prediction of IPV recidivism. These mixed findings may well reflect the fact that the association between alcohol misuse and IPV recidivism is complicated. According to the results reported by Cattaneo and Goodman (2005), the association between substance misuse and IPV has been measured in various ways with varying results, in line with the results found by Murphy et al. (2003). Snow Jones and Gondolf (2001) reported that perpetrators' histories of substance abuse was not a significant predictor of reabuse, but recent alcohol use during the follow-up period was an important indicator. Although the PAPS total score related significantly with all recidivism outcomes, there exists only one published study on the predictive validity of this new instrument to our knowledge. As a consequence, the PAPS has not outgrown the pilot phase, and we agree that "practice applications should await further validation" (Murphy et al., 2003, p. 1101).

Actuarial Risk Assessment Measures—General Violence Measures

The general violence risk assessment literature has advanced at a rapid pace in the past two or three decades (e.g., Monahan et al., 2001; Otto & Douglas, 2010). Those developments, in combination with the importance of antisocial personality disorder characteristics (Hilton et al., 2001) and the extent to which many predictors of criminal behavior and violence appear to translate across settings and populations (e.g., Bonta, Law, & Hanson, 1998), have led several researchers to examine the extent to which general violence risk assessment measures predict IPV.

Our review identified several published studies that have evaluated the capacity of three nondiscretionary measures that were not developed with the intention of evaluating IPV risk specifically to predict recidivism among male abusers. Two published papers reported results from the VRAG (Harris et al., 1993; Quinsey et al., 2006), three studies included data on the PCL-R[11] (Hare, 1991, 2003), and four studies reported results for the Level of Service Inventory-Revised (LSI-R) and LSI-OR (Andrews & Bonta, 1995, 2000, 2001; Andrews, Bonta, & Wormith, 1995).

The Violence Risk Appraisal Guide (Harris et al., 1993; Quinsey et al., 2006). The VRAG is an actuarial instrument intended to assess the risk of violent criminal recidivism in offenders (Harris et al., 1993); it was not developed with IPV in mind, specifically. The instrument initially was constructed and validated in a prospective

study with a sample of 618 males accused of a serious criminal offense and referred to a maximum security psychiatric hospital for pretrial or presentence assessment (Harris et al., 1993). For each of the participants, the researchers had information regarding 50 predictor values that had previously been shown to empirically relate to violent or criminal behavior (Quinsey et al., 2006). The variables pertained to the following categories: sociodemographic information, childhood problems, adult adjustment, index offense characteristics, and psychological assessment variables (Quinsey et al., 2006). A bivariate analysis was used to compare differences on the predictor variables between individuals who did and did not recidivate (Quinsey et al., 2006). Only significant variables were considered for inclusion on the instrument, and in cases where variables were highly correlated with each other, only the variables with the highest correlation were considered for further analysis. A stepwise multiple regression analysis determined which of the variables independently added to the prediction of violent recidivism. Twelve items were ultimately selected for inclusion in the actuarial instrument, and further analysis led to the development of a weighted scoring system. VRAG scores can range from -26 to $+38$, with a mean score of approximately zero (Quinsey et al., 2006).

To score the VRAG, one needs access to a comprehensive psychosocial history including a breadth of information including variables such as conduct during childhood, family problems, criminal history, and present psychological problems (Rice et al., 2010). As collateral information is needed to properly score the VRAG, clinically compiled records often are used (Rice et al., 2010). The authors recommend users to undergo training and exhibit proficiency with the VRAG as well as the PCL-R (an item on the VRAG; Rice et al., 2010). As discussed in the following text, considerable training and expertise is required to code the PCL-R, making the VRAG less accessible to professionals without advanced training and degrees. In our review of the literature, two published studies were identified that used the VRAG to assess the likelihood that a group of offenders with a history of IPV would recidivate violently (Grann & Wedin, 2002; Hilton et al., 2001).

Under the supposition that risk factors for general violent behavior would also predict the specific recidivism of wife assaulters, Hilton et al. (2001) tested the capacity of the VRAG to predict recidivism among violent offenders and wife assaulters. The sample consisted of 508 violent offenders of which 88 offenders committed acts of violence against their spouse.[12] The VRAG was retrospectively coded and had excellent IRR (ICC > .80; details on IRR can be found in previous reports [Harris, Rice, & Cormier, 1991; Quinsey et al., 2006; Rice, Harris, Lang, & Bell, 1990]). The outcome measure violent recidivism "was defined as a new criminal charge for an offense against a person or a readmission to a psychiatric facility for violent behavior that could have resulted in such a criminal charge" (Hilton et al., 2001, p. 414). The recidivism data was obtained retrospectively from file information. Of the wife assaulters, 80 (90.9%) men had the opportunity to recidivate during the mean follow-up period of almost seven years ($M = 82.5$ months, $SD = 56.0$ months). Nineteen (23.8%) of the 80 men who were at risk recidivated violently. The remaining violent offenders

consisted of 388 men who had the opportunity to recidivate during the mean follow-up period of almost six years (M = 69.3 months, SD = 59.5 months); of these 388 men, 172 (44.3%) recidivated violently.

The authors (Hilton et al., 2001) compared the general violent offenders and wife assaulters on the VRAG and PCL-R (Hare, 1991). The wife assaulters were significantly different from the violent offenders in terms of their PCL-R total scores and all other VRAG variables, with the exception of alcohol abuse. In general, the wife assaulters had lower risk scores, although they were more likely to have been suicidal (suicidal motive: wife assaulters, 11.4%; other offenders, 2.4%) or jealous (jealous motive: wife assaulters, 43.2%; other offenders, 5.0%) during the index offense.

The predictive accuracy of the VRAG regarding violent recidivism of the wife assaulters (note: this was not necessarily wife abuse) was $r(79)$ = .42, $p < .001$, and AUC = .75 (SE = .07). The authors also conducted a survival analysis to take into consideration the difference in follow-up periods between the two groups of offenders. There were no significant differences found in survival patterns or in the association of the VRAG with violent recidivism between offenders with more or less follow-up time among the general violent offenders ($r[222]$ = .50 vs. $r[199]$ = .39, z = 1.40, ns) nor in the wife assaulter group ($r[42]$ = .32 vs. $r[34]$ = .58, z = 1.40, ns). Unfortunately, this study does not report any results on the prediction of IPV recidivism, specifically.

A retrospective study in Sweden focusing on the SARA (Kropp et al., 1995, 1999) also included a small amount of information on the VRAG. Grann and Wedin (2002; study discussed in the following text) included the VRAG to obtain concurrent validation data for the SARA with general risk assessment instruments. The scores on the VRAG ranged from −14 to 28 (M = 5.47, SD = 7.94). The SARA total score correlated significantly with the VRAG total score (r = .33, $p < .01$). The correlation between the VRAG and SARA Part 1 scores reflected a stronger correlation (r = .49, $p < .01$) compared to the correlation between VRAG and the scores of Part 2 of the SARA (r = −.01, ns). This is to be expected because the Part 1 items from the SARA measure the general criminal history, whereas the Part 2 items of the SARA assess spousal assault history, specifically.

With only two published studies available and an absence of any published literature examining the VRAG's capacity to predict future IPV specifically (i.e., as opposed to general violence/recidivism, as was the intention of the developers), there is currently little evidence to support clinicians in adopting this measure for use in cases involving IPV (see Table 3).

The Psychopathy Checklist-Revised (Hare, 1991, 2003). The PCL-R (Hare, 1991, 2003) is a 20-item rating scale intended to assess the construct of psychopathy. Items are rated on a 3-point ordinal scale (0, 1, and 2) and total scores range from 0 to 40. According to the author, the PCL-R score provides a dimensional score of the extent to which an individual matches the criteria of a "prototypical psychopath" (Hare, 2003, p. 17). Factor 1 is composed of Facet 1 (interpersonal [five items]) and Facet 2 (affective [four items]), and Factor 2 is composed of Facet 3 (lifestyle [five items]) and Facet 4

TABLE 3. Psychopathy Checklist – Revised (PCL-R) & Violence Risk Appraisal Guide (VRAG)

Study	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Hilton et al. (2001) *Canada* PCL-R	Subsample: Offenders identified as wife assaulters N = 88	37 yrs	Retrospective	New criminal offence for a crime against a person or readmission to a psychiatric facility for violent behaviour	23.8%	IRR: $r > .80$ $K > .70$	Not reported	PCL-R & violent recidivism: $r(74) = .39$, $p < .001$ PCL-R total: general violent recidivists ($M = 18.24$) and wife assaulters ($M = 11.92$) ($X^2 = 5.80$, $p < .001$).
VRAG							Not reported	Information from the sub sample of wife assaulters: VRAG & violent recidivism: • $r(79) = .42$, $p < .001$ • $AUC = .75$

Study	Sample	Age	Design	Outcome	Recidivism rate	Reliability	Validity	Predictive validity
Grann & Wedin (2002) *Sweden* **PCL-R**	Male offenders N = 88	35 yrs	Retrospective file review (7.8 yrs)	Any reconviction of a hands-off or hands-on violent and/or sexual behavior, threats, or violation of no-contact orders	6 months = 15% 1 year = 18% 2 years = 24% 5 years = 25%	Total score: N = 18 ICC = .85 Indiv. items: *Mean* K = .58, (range = .30 – 1.0)	SARA: r = .59, p < .01	Violent recidivism w/in 1 yr after release PCL-R AUC = .71, 95% CI [.60, .80]
VRAG							SARA: r = .33, p < .01	Violent recidivism w/in 1 yr after release: • VRAG AUC = .75, 95% CI [.65, .84]
Hilton et al. (2008) *Canada* **PCL-R**	Sample 1: Male offenders with a history of IPV N = 303	S1: 36 yrs S2: 35 yrs	Retrospective follow-up (5.0 yrs)	Any subsequent violent assault against an (ex) wife or (ex) common-law wife	S1 = 49% S2 = 41%	Not reported	VRAG: r = .72* ODARA: r = .55* SARA: r = .55* DA: r = .36* DVSI: r = .34* DVRAG: r = .72* *p < .001	PCL-R Recidivism: • r = .22, p < .001 • AUC = .66, p < .001
VRAG	Sample 2 : Male offenders with a history of IPV N = 346					Not reported	PCL-R: r = .72* ODARA: r = .53* SARA: r = .43* DA: r = .21* DVSI: r = .31* DVRAG: r = .66* *p < .001	VRAG Recidivism: • r = .19, p < .001 • AUC = .67, p < .001

Note. IRR = inter-rater reliability; AUC = area under the curve; ICC = intraclass coefficient; CI = confidence interval; IPV = Intimate partner violence; Indiv. = individual; DA = Danger Assessment Guide; DVRAG = Domestic Violence Risk Appraisal Guide; DVSI = Domestic Violence Screening Instrument; ODARA = Ontario Domestic Assault Risk Assessment; SARA = Spousal Assault Risk Assessment.

(antisocial [five items]). The PCL-R generally is coded after a detailed semistructured clinical interview is completed with the client and a thorough review of case history information, collateral reports, and official records whenever possible. Up to five items can be omitted, and the total score is then prorated (see Hare, 2003, pp. 6–7). Hare (2003) cautions that given the new four-facet model with fewer items per facet, "The score on each facet should be based on all of the items in the facet whenever possible" (p. 6). Given the dramatic and life-altering implications of being labelled a psychopath, Hare (2008) recommends the PCL-R be used with strict adherence to ethical guidelines. In addition, the *Hare Psychopathy Checklist-Revised Manual* (Hare, 2003) recommends that for clinical use, the clinicians who use or supervise the use of the PCL-R have an advanced degree; have completed courses in statistics, psychopathology, and psychometric theory; and have knowledge of the psychopathy literature—the relevant qualifications and adequate training and experience with the PCL-R. Although not intended originally as a violence risk assessment measure, the PCL-R consistently has been found to be moderately correlated, in the expected direction, with various antisocial and criminal behaviors as well as treatment outcomes in diverse samples. Research findings span the past three decades and include an impressive amount of laboratory and field research with civil, forensic, and correctional samples (Herve & Yuille, 2007). Of particular relevance to the present discussion, research examining subgroups of batterers suggests that forensic assessments, such as the PCL-R, might add invaluable information when evaluating and treating IPV offenders, particularly generally violent batterers (i.e., men who offend both within the home and also engage in antisocial behavior more generally within the community; Spidel et al., 2007).

Although some experts have suggested that abuse within the family reflects unique predictors (e.g., Kropp & Hart, 2000), Hilton et al. (2001) proposed that a consideration of longitudinal research (see Magdol, Moffit, Caspi, & Silva, 1998) points to early antisocial behavior as a robust indicator of abuse in intimate relationships. In addition, the strength and consistency of the body of research revealing psychopathy to be a central variable in general violence risk assessment is firmly established in the literature; thus, it stands to reason that psychopathic characteristics could be an essential consideration for assessors in the IPV field as well. The relevance of this malignant personality disorder to criminal and antisocial behavior has resulted in leading authorities including it as an item on existing general risk assessment tools (VRAG and HCR-20; Hare, 2008). Our review of the literature identified three published studies of relevance to the present discussion.

Hilton et al's. (2001) retrospective study of 508 offenders included a subsample of 88 male wife assaulters. They found that partner abusers received lower mean PCL-R scores (11.92, $SD = 9.06$) than other offenders (18.24, $SD = 9.03$). A strength of their study was that the authors made use of multiple and diverse sources of information to determine recidivism (e.g., information obtained from parole, corrections, coroner's office, readmission to hospital for violence). Limitations included the extent to which IPV reabuse may not be accurately reflected on official data sources and the absence

of self-report or victim reports. Most importantly, the outcome was not specific to IPV recidivism but general violent offending. As noted earlier, of the IPV offenders, 80 (91.0%) had an opportunity to reoffend, with an average time at risk of nearly seven years. Nineteen of the wife assaulters recidivated violently (23.8%); interestingly, this was substantially lower than the other offenders (44.0%). The PCL-R score was inversely correlated with suicidality and jealousy, and neither of those variables was a significant predictor of wife abusers recidivating. The association between the PCL-R and violent recidivism for the wife assaulters was moderate and significant ($r[74] = .39, p < .001$). The authors concluded that violent recidivism by serious IPV offenders reflects the same variables that are predictive of violent reoffending among general offenders. The generalizability of this study is limited by the fact that the offenders were patients in a maximum security hospital; however, when the authors removed the subgroup of participants with a diagnosis of schizophrenia (*Diagnostic and Statistical Manual of Mental Disorders, Third Edition* [*DSM-III*] criteria = 13.6%), the results remained stable. In addition, the index offenses were more serious than would be found in an unselected cohort of IPV perpetrators. More than half of the men had an index offense for murder or manslaughter (59.1%), thereby ruling out the possibility of offending against the same victim. Finally, as noted previously, the outcome criteria required that the reoffense had necessarily come to the attention of the criminal justice system.

The second study including PCL-R data we found was a retrospective study of spousal abuse in Sweden we have mentioned previously. Grann and Wedin (2002) sampled 88 male batterers who had been court-ordered to undergo forensic psychiatric evaluations between 1988 and 1990. In total, 25 (28%) of the men perpetrated IPV assaults for which they were convicted during follow-up (5 years, 7 months). The PCL-R scores were predictive of recidivism within 1 year after release (AUC = .71, 95% CI [.60–20]); Factor 1 (AUC = .70, 95% CI [.59–.79]) and Factor 2 (AUC = .71, 95% CI [.60–.80]). The authors did not report the predictive validity for the other follow-up periods. As the authors noted, their study suffered because of a reliance on files to code the measures (reflected in several cases in which items had to be omitted because of insufficient information). Grann and Wedin (2002) suggested that the poor measurement reliability may have resulted in restricted effect sizes in the risk ratios and AUCs. A reliance on official records at follow-up was also noted to have been a study limitation. Compared to several of the other studies examining IPV risk assessment with general violence risk assessment tools, this study had the particular strength of using spousal assault recidivism as the outcome criterion, as opposed to any recidivism/violent recidivism.

The third and final study identified in our review was published by the authors of the ODARA who were seeking a means of making the instrument a more comprehensive assessment of IPV risk (Hilton, Harris, Rice, et al., 2008). They used two samples of male offenders with a history of IPV; one group was a sample of 303 men from the construction and validation of the ODARA (Hilton et al., 2004), and the other sample was made up of 346 men who similarly had police records and IPV

histories. Using records, the authors scored the VRAG, PCL-R, SARA, DA, and DVSI to see if they could add a more in-depth picture of risk to the ODARA. For Sample 1, all of the instruments had small but significant correlations ranging between $r = .12$ and $r = .22$, with dichotomous wife assault recidivism; the PCL-R had the highest correlation ($p < .001$). Overall, the PCL-R performed best with further bivariate correlations between the PCL-R and the number of recidivistic incidents, victim injury in recidivism, number of severe incidents, and Cormier-Lang recidivism scores as follows: .28, .31, .26, and .23 (all $p < .001$). In the receiver operating characteristic (ROC) analysis, the PCL-R was among the strongest predictors of dichotomous IPV recidivism (AUC = .66), with only the VRAG exhibiting a slightly larger AUC (.67). Because of the performance the PCL-R exhibited in this study, the authors paired the PCL-R with the ODARA to result in the DVRAG, a more in-depth approach to risk assessment (Hilton, Harris, Rice, et al., 2008).

Both theoretical and empirical literature support the inclusion of psychopathy as a valuable source of information in IPV assessments (Hilton et al., 2001; Spidel et al., 2007). We would strongly recommend that although the PCL-R is an appropriate consideration in risk assessments for spousal violence, it is just one indication of IPV risk, and thus should not be the only measure used to inform a risk assessment for partner abuse and/or femicide. To clarify, there are many men who abuse their partners who would not necessarily receive high scores on a measure of psychopathy; and similarly, there are many variables other than psychopathy that are positively and moderately correlated with partner abuse (see Table 3).

Level of Service Inventory-Revised (Andrews & Bonta, 1995, 2000, 2001; Level of Service Inventory-Ontario Revision, Andrews et al.*, *1995). The LSI (Andrews & Bonta, 1995, 2000, 2001) is a measure of risk for general offending that had its impetus in the groundbreaking work of Drs. Don Andrews and James Bonta, two leading Canadian authorities in the field of corrections and mental health. The LSI was developed as an assessment of general criminal recidivism and originally was intended to inform probation and parole supervision (Girard & Wormith, 2004). The LSI now has several derivations including the LSI-OR (Andrews et al., 1995). Andrews, Bonta, and Wormith (2010b) noted that the Level of Service assessment measures draw considerably from their work on the risk-need-responsivity (RNR) model (Andrews, Bonta, & Hoge, 1990). The authors also indicated that three major sources of information have informed the development of this family of measures: input from correctional and forensic professionals, broad theory of human behavior and criminal behavior, and research on the prevention and assessment of criminal behavior (Andrews et al., 2010a).

The LSI-R (Andrews & Bonta, 1995, 2000, 2001) consists of 54 indicators of risk/need under 10 subheadings: criminal history (10 items), education/employment (10 items), financial (2 items), family/marital (4 items), accommodation (3 items), leisure/recreation (2 items), companions (5 items), alcohol/drug problem (9 items), emotional/personal (5 items), and attitude/orientation (4 items; see Andrews et al., 2010b). Each item is

coded dichotomously (1 [*present*], 0 [*absent*]). If there is inadequate information, every effort should be made to obtain the information; however, the authors indicate the LSI-R can be used effectively with up to five omitted items (Andrews & Bonta, 2001). Subcomponent and total scores are calculated by summing the number of checked items. Some of the dynamic items include a more precise 4-item Likert rating scale inviting assessors to indicate *a very unsatisfactory situation* (0) *to a very satisfactory situation* (3). Zero and 1 ratings are considered "checked" items, and scores of 2 and 3 are considered "not checked" (Andrews et al., 2010b). The LSI-R (Andrews & Bonta, 2001) and LSI-OR (Girard & Wormith, 2004) both provide for a clinical override. As the authors describe, every effort has been made to provide a comprehensive measure; however, it is impossible to foresee all possible scenarios. Therefore, trained professionals are encouraged to note when unique circumstances require attention and when the quantitative risk/needs assessment should be revised to reflect professional discretion. According to the authors, the professional override is used relatively rarely (less than 1 in 10 assessments, with fewer than 5% increasing the risk/need level and fewer than 5% decreasing the risk/need level; Andrews & Bonta, 2001). In spite of the allowance for professionals to override the overall risk determination, the LSI is categorized as an actuarial measure because when scoring the instrument, contingency tables are used that link scores to outcomes (Andrews, Bonta, & Wormith, 2010a).

The LSI-OR was intended to address perceived inadequacies in the LSI from the perspective of case managers and correctional staff and to provide a common risk/needs measure that could be used to facilitate the continuity of care from institutions to community (Girard & Wormith, 2004). Girard and Wormith (2004) described the LSI-OR as a "theoretically and empirically developed risk/need assessment instrument designed to predict an offender's risk of reoffending" (p. 155). Items on both versions of the LSI are preferably scored from a combination of data sources including file reviews, official records, and interviews with the client and collaterals (e.g., family member; Andrews et al., 2010b). Detailed manuals and training are available and allow for the LSI to be scored by individuals without formal psychological assessment training, although the authors caution that a basic understanding of the principles of psychological test interpretation is essential (Andrews et al., 2010b). For a more detailed discussion of the various forms of the LSI and its development, see Andrews et al. (2010b) and Girard and Wormith (2004).

Our review of the literature revealed four studies that employed the LSI-R or LSI-OR and examined predictive validity with IPV offenders (see Table 4). The first, a prospective validation study (Girard & Wormith, 2004), evaluated the long-term predictive accuracy of the LSI-OR in adult male inmates ($N = 454$) and community probationers ($N = 176$). Recidivism data were obtained from an automated Offender Management System (OMS) and from a national police database, the Canadian Police Information Centre (CPIC). Follow-up took place over a 2.5-year period. Overall, 24.1% ($n = 152$) of the sample recidivated during the follow-up and had at least one violent conviction, with the institutional group being more likely to recidivate violently than the probationers (27.1% vs.16.5%, $\chi^2[1, n = 152] = 58.13, p < .001$).

TABLE 4. Level of Service Inventory – Revised (LSI-R)/ Level of Service Inventory – Ontario Revision (LSI-OR)

Study	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Girard & Wormith (2004) *Canada* LSI-OR	Male perpetrators of IPV *N* = 630	32 yrs	Prospective validation study (up to 2.5 yrs)	Conviction for any offence	54.4% of male offenders recidivated with at least one conviction 24.1% recidivated with at least one violent conviction *Note*. IPV specific offending not reported.	IRR: *K* = .58 Internal consistency of 43 General Risk/ Need items: • α = .91 Internal consistency of Specific Risk/Need Section: • α = .62 Test-retest at 1 month: General Risk/ Need items = .88 (*p* < .001)	Not reported	IPV offenders recidivated* violently at significantly higher rates than other offenders including the violent (non-IPV) group (*t* (200.48) = −4.34, *p* < .001) IPV of fenders (*M* = 21.44, 95% CI [19.94, 22.94]) had significantly higher General Risk/Need scores than other offenders including the violent (non-IPV) group (*M* = 19.77, 95% CI = 18.99 to 20.54; *t* (628) = −1.99, *p* < .05)

| Hanson & Wallace-Capretta (2004) *Canada* *LSI-R* | Male perpetrators of IPV $N = 320$ | 35 yrs | Prospective (5 years) | Charge or conviction for any offence | 17.2% recidivated with a violent offence 25.6% recidivated with any offence | Not reported | Not reported | *Note. This is not IPV specific violent recidivism. Correlations: • Violent recidivism: $r = .32$, $p < .001$ • Any recidivism: $r = .40$, $p < .001$ AUCs: • Violent recidivism $= .73$ $(SD = .04)$ • Any recidivism $= .76$ $(SD = .03)$ |

(continued)

TABLE 4. (continued)

Study	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Hendricks et al. (2006) USA LSI-R	Male perpetrators of IPV $N = 200$	Not reported	Prospective cohort study	Any offense for domestic violence to an intimate partner	10.6% Of participants who completed program 38.8% of participants who did not complete program	Not reported	Not reported	Overall classification rate when Risk & Need Scales entered independently into logistic regression: = 64% Overall classification rate LSI-R total score = 66% • Specificity = 67% • Sensitivity = 60%
Hilton et al. (2010) Canada LSI-OR	$N = 150$ incarcerated IPV offenders	30 yrs	Retrospective follow-up (8.0 yrs)	Documented IPV recidivism (police report or official criminal record)	27%	Not reported	Not reported	IPV AUC = .50, 95% CI [.40, .61]

Note. IRR = inter-rater reliability; AUC = area under the curve; ICC = intraclass coefficient; CI = confidence interval; IPV = Intimate partner violence.

Although the general recidivism rates were very similar, the domestic violence offenders recidivated violently at significantly higher rates than the other groups including the violent (nondomestic) group (t[200.48] $= -4.34$, $p < .001$); however, the study did not report on IPV offenses, specifically. The domestic violence group had significantly higher general risk/need scores than the nondomestic violent offenders ($M = 21.44$, $SD = 9.06$, CI [19.94–22.94] vs. $M = 19.77$, $SD = 8.70$, CI [18.99–20.54]; t[628] $= -1.99$, $p < .05$). Kappa coefficients of agreement[13] between raters was fair (.58). The internal consistency of the 43 general risk/need items was high with an α of .91. Internal consistency was not as strong for the specific risk/need section with an α coefficient of .62. The 1-month test–retest reliability was .88 ($p < .001$) for the general risk/need section ($n = 18$) but only .12 ($p = ns$) for the specific risk/need section. In terms of predictive accuracy, the general risk/need total scores were slightly stronger predictors for the nondomestic violence offenders for general recidivism, but when the outcome was focused on violent recidivism, the LSI-OR had similar predictive accuracy in the domestic violent offenders than in the nondomestic violent offenders ($\gamma = .44$).

The second study, by Hanson and Wallace-Capretta (2004), had the strength of employing a prospective design in a relatively large sample of male offenders recruited voluntarily from five community treatment programs across Canada ($N = 320$). The study strengths also included collecting collateral information from the men's partners and reporting official criminal recidivism across a long follow-up period (39–73 months [$M = 58$; $SD = 7.7$]). In addition, the authors explored diverse indicators of violence risk in addition to the LSI-R, including for instance, a measure to evaluate abuse in the family origin, a marital adjustment scale, and a consumer satisfaction scale. The study's implications for the field are limited because of the failure to report IPV recidivism separately from general and violence reoffending. The recidivism information including both charges and convictions was obtained from national police records; however, the authors noted that the identity of the victims was unavailable, making it impossible to ascertain if the offenses pertained to IPV or not. Regardless, the findings demonstrated that many of the same variables that are associated with recidivism risk are relevant to male IPV offenders. Correlation coefficients for violent recidivism (IPV) $= .32$ ($p < .001$) and general recidivism $= .40$ ($p < .001$) as well as ROC AUC (violent recidivism [IPV] AUC $= .73$ [$SD = 0.04$]; general recidivism AUC $= .76$ [$SD = 0.03$]) were moderate. It is also noteworthy that the adaptations to the LSI-R to make it more relevant to recidivism risk among abusive men were only partially successful. For instance, the original Criminal Associates Scale showed stronger predictive accuracy than the revised questions that queried an association with abusive peers. Similarly, the partners' reports on the men's substance abuse did not improve on the predictive accuracy provided by the men's substance abuse self-reports.

The most recent study to report on the LSI measures in relation to IPV was published by Hilton et al. (2010) who sampled 150 male inmates who were eligible for a domestic violence program. The authors asserted that to their knowledge, this study

represented the first validation of the ODARA in an incarcerated sample and only the second test of any domestic violence risk assessment measure among inmates. The authors suggested that because of the high rate of IPV histories among inmates, this is a particularly important gap in research. Given that the central purpose of the study was to determine if the ODARA could predict recidivism in incarcerated male IPV offenders, this study will also be examined in relation to that measure. For present purposes, it should be noted that the LSI-OR was completed as part of the regular intake procedure and recorded for the study. The mean LSI-OR score ($n = 140$) was 31.04 ($SD = 6.08$). The ODARA (AUC = .64) was found to be a stronger predictor of domestic violence recidivism than the LSI-OR (AUC = .50). Specifically, the authors reported that when "total LSI-OR score was entered into the first block of a binary logistic regression to predict dichotomous domestic violence recidivism, ODARA score was selected in a forward conditional test ($\alpha = .05$, final model Nagelkerke $R^2 = .04$), but not vice versa" (p. 824). The LSI-OR (AUC = .62) performed better than the ODARA (AUC = .59) in predicting violence when the victim had an unknown relationship with the offender, and the ODARA (AUC = .69) and LSI-OR (AUC = .70) predicted the occurrence of any postrelease charges equally. In contrast to the Hanson and Wallace-Capretta (2004) article discussed previously, neither the LSI-OR (AUC = .58) nor the ODARA (AUC = .56) were strongly associated with overall violent recidivism. Correlations with other treatment variables are available in the article. The study limitations included scoring the ODARA from institutional file reviews completed by research assistants rather than by institutional staff, suggesting that future research should explore the use of that measure when coded by staff. Although it should be noted that the study reported excellent IRR. The research assistants and the first author coded 10 training cases (ICC = .95) and IRR at follow-up on 10 cases (ICC = .94) for time at risk and ICC = .76 for recidivism. Discrepancies were resolved by consensus.

In sum, the LSI-OR has been found to be significantly associated with general criminal recidivism among former inmates but had a small, nonsignificant association with domestic violence recidivism, specifically (Hilton et al., 2010). The LSI has demonstrated small predictive effects for domestic violence in community samples (Hendricks, Werner, Shipway, & Turinetti, 2006). Despite the common perception that specialized risk assessments may be superior for subgroups of offenders, the LSI-OR results suggest that a generalized tool might be a reasonable means of assessing exceptional groups of offenders (i.e., IPV, sexual offenders, mentally ill offenders), although further research that examines IPV reoffending specifically is required (see Girard & Wormith, 2004).

Structured Professional Judgment—Intimate Partner Violence Specific Measures

Spousal Assault Risk Assessment Guide (Kropp, Hart, Webster, & Eaves, 1994, 1995, 1999, 2008). The SARA is a SPJ measure, developed as an inventory of risk factors for IPV rather than a psychological test, thereby allowing for use by a wide

range of professionals such as lawyers, correctional staff, mental health professionals, and victim's advocates (Kropp et al., 2008). The SARA was developed by several leading authorities in forensic mental health, including Drs. Kropp, Hart, Webster, and Eaves who have coauthored many similar SPJ guides and measures to inform violence risk assessments. The authors first thoroughly reviewed the clinical and empirical literature pertaining to risk for violence, and more specifically spousal violence (Kropp & Gibas, 2010). Risk factors were included in the instrument if they empirically distinguished between individuals who abused their partners and those who did not, the risk factor was associated with spousal assault recidivism, and/or the risk factor was included in previous professionally accepted guidelines (Kropp & Gibas, 2010). The 20 items that make up the SARA each fall into one of the following categories: criminal history, psychosocial adjustment, spousal assault history, index offense, and other considerations (Kropp et al., 2008). The items are scored using a 3-point nominal scale: N (*no or absent*), P (*the item possibly or partially applies*), and Y (*yes or present*; Kropp et al., 2008). Based on the scoring of the 20 items, the rater makes a summary risk judgment indicating that the individual is at low, moderate, or high risk for committing future IPV (Kropp et al., 2008). Although evidence suggests a general linear relationship between SARA assessment total scores and IPV-related outcomes, the SARA is not an actuarial measure. The SARA authors recommend that users be aware that individual risk factors, if present, may alone indicate high risk in the absence of a high total score (e.g., pathological jealousy and access to a weapon; Kropp et al., 1999).

Although the SARA can be used by individuals who are not mental health professionals, the authors note that it is necessary that the file contain a report by a mental health professional such as a psychologist or a psychiatrist because there are items on the measure that relate to the accused/perpetrator's current and past mental health status (Kropp et al., 2008). Finally, the SARA can be coded from file (i.e., in the absence of an interview) but that being said, it is always preferable to triangulate your data collection and obtain information from as many sources as possible (e.g., victim interview, perpetrator interview, interviews with children and other collaterals) to supplement file reviews. Information needed to score this measure includes collateral reports from the victim or witnesses, a criminal history, information regarding past assaults, and relationship history (Kropp et al., 2008). The SARA is most commonly used in matters such as pretrial release, sentencing, correctional intake and discharge, and to inform decisions concerning the duty to warn (Kropp et al., 2008).

Of the seven studies we found that discussed the psychometric properties of the SARA, two were conducted in Sweden (Belfrage et al., 2011; Grann & Wedin, 2002), two in the United States (Heckert & Gondolf, 2004; Williams & Houghton, 2004), and three in Canada (Hilton, Harris, Rice, et al., 2008; Hilton et al., 2004; Kropp & Hart, 2000; see Table 5). This IPV risk assessment measure has been validated predominantly in samples of male offenders and probationers convicted of IPV-related offenses (e.g., Belfrage et al., 2011; Kropp & Hart, 2000; Williams & Houghton, 2004). The research includes studies using SARA assessments completed by various

TABLE 5. Spousal Assault Risk Assessment (SARA)

Study/Administration	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/Convergent Validity	Predictive/Postdictive Validity
Kropp & Hart (2000) *Canada* Used recommended SARA administration	Sample 1: Male probationers $N = 1671$ Sample 2: Male inmates $N = 1010$ Recidivists: $N = 52$ Nonrecidivists: $N = 50$	*Mdn*: 32 *yrs*	Postdictive (NA)	Severe IPV recidivism defined as: threatened to kill, threatened with weapon, kicked, bit, or hit with fist, hit or tried to hit with something, forced sex, choke/strangle, beat up, used weapon;	NA	IRR: Individual items scored on 3 pt scale: Mdn ICC = 0.65; range = .45–86. Total score: ICC = .84* Critical items: ICC = .22* Low vs. moderate vs. high summary risk rating: ICC = .63* * $p < .05$ Item homogeneity: MIC = 0.15 Internal consistency: $\alpha = 0.78$	Total score: • PCL:SV = .43 (p <0.001) • GSIR = −0.07 (ns) • VRAG = 0.29 (ns)	Total SARA scores did not significantly differ between recidivists (17.69) and non recidivists (15.68) ($p < .068$) Summary risk rating: • AUC= .70 (*SE* = .06)

						IRR ($N = 18$)	Total score:	AUCs for SARA
Grann & Wedin (2002) *Sweden* Did not use recommended SARA administration	Male offenders $N = 88$	35 yrs	Retrospective file review (7.8 yrs)	Any reconviction of a hands-off or hands-on violent and/or sexual behavior, threats, or violation of no-contact orders	6 months = 15% 1 year = 8% 2 years = 24% 5 years = 25%	Total score: • ICC = .85 ($p < .01$) For indiv. items, mean k = .58, (range .30–1.0)	• VRAG: r = .33 ($p < .01$) • PCL-R: r = .59 ($p < .01$)	total score and IPV recidivism within: • 6 months = .52, 95% CI [.41, .63] • 1 yr = .59, 95% CI [.48, .70] • 2 yrs = .63, 95% CI [.51, .73] • 5 yrs = .65, 95% CI [.5, 1.77] Correlations: • # prior convictions: r = .39* • # prior violent convictions: r = .29* * ($p < .01$)

(continued)

TABLE 5. (continued)

Study/ Administration	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Heckert & Gondolf (2004) USA Did not use recommended SARA administration	Male batterers admitted to batterer treatment groups & their partners N = 499	Not reported	Prospective (3 month follow-ups for a period of 15 months)	Multiple outcomes: • Repeat reassault • One time reassault • Threatening reassault • Controlling behaviour; or verbal abuse	Repeat re-assaulters: 23% One-time re-assaulters: 12% Threatening reassault: 20% Controlling behavior or verbal abus-ers: 26% No abuse: 19%	Not reported	Not reported	Simulated SARA total scores: • AUC = .64; CI not reported • Sensitivity = 43% • False positives = 27%
Hilton et al. (2004) Canada Did not use recommended SARA administration	Male offenders N = 589 + 100 cases for cross validation	38 yrs	Retrospective follow-up (4.8 yrs)	Any subsequent violent assault against an (ex-) wife or (ex-) common-law wife	Construction sample: • 30% Cross valida-tion sample: • 26%	Not reported	ODARA: r = .60* DA: r = .43* DVSI: r = .53* *p < .01	Construction sample: • AUC = .64, 95% [CI ±.05] Cross Validation: • AUC = .54, 95% [CI ±.05]

| William & Houghton (2004) *USA* Used recommended SARA administration | Sample 1: Male offenders arrested for IPV against female partners *N* = 1465 Sample 2: Female partners of male IPV perpetrators *N* = 125 | 32 yrs | Prospective validation (S1: 18 month follow-ups) (S2: 6 month follow-ups) | Arrests for violations of domestic violence restraining orders and for domestic violence reoffending | 29% | Not reported | Total score: • DVSI: *r* = .54 Summary risk estimate: • DVSI: *r* = .57 | • AUC = .65, *p* < .000 |

(continued)

TABLE 5. (continued)

Study/ Administration	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Belfrage et al. (2011) *Sweden* Used recommended SARA administration	Male IPV perpetrators N = 429	39 yrs	Prospective (18 month follow-up)	Actual, attempted, or threatened physical or sexual violence against a past or current intimate partner	21% had further contact with police for incidents related to IPV	Stability of total scores between first and second police contacts: ICC = .76	Not reported	Mean total score: • Recidivists = 11.45 (SD = 6.10), • Non-recidivists 8.86 (SD = 5.84) • (t(92) = 3.79, p < .001) Total scores: • r = .18, p < .001 • AUC = .63 Summary risk ratings: • r_{pb} = .09, p = .06 • AUC = .57 (SE = .03)

Note. IRR = inter-rater reliability. AUC = area under the curve. ICC = intraclass coefficient. MIC = mean inter-item correlation. CI = confidence interval. IPV = intimate partner violence; DA = Danger Assessment Scale; DVRAG = Domestic Violence Risk Appraisal Guide; DVSI = Domestic Violence Screening Instrument; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist – Revised; VRAG = Violence Risk Appraisal Guide.

professionals, including police officers (Belfrage et al., 2011), correctional staff (Kropp, & Hart, 2000), and graduate level researchers (e.g., Hilton et al., 2004).

Compliance with the assessment procedures set out in the manual by the SARA authors has been very mixed in the cross-validation research to date. Two of the studies completed the SARA assessments as recommended in the manual using a combination of file review and interviews (Kropp & Hart, 2000; William & Houghton, 2004), and four of the studies coded the SARA from files only (Grann & Wedin, 2002; Hilton, Harris, Rice, et al., 2008; Hilton et al., 2004). Belfrage and colleagues (2011) used SARA assessments completed by police officers in the course of their investigations. Heckert and Gondolf (2004) employed a simulated version of the instrument that contained only 16 of the measure's 20 items, 10 of which were the same as the original instrument and six were noted to be similar to the actual SARA items.

Reliability statistics were provided in just two of the SARA studies identified in our review. The IRR statistics reported varied considerably depending on the component of the measure of interest. Total scores were found to have excellent interrater agreement (Cicchetti & Sparrow, 1981) across both studies, with ICCs of .84 ($N = 86$, $p < .05$; Kropp & Hart, 2000) and .85 ($N = 18$, $p < .01$; Grann & Wedin, 2002). Kropp and Hart (2000) also found that the IRR for Part 1 (.68, $p < .05$) and Part 2 (.87, $p < .05$) were good to excellent. The IRR for the summary risk ratings of low versus moderate versus high was good (ICC = .63, $p < .05$). The IRR for critical items was poor (ICC = .22, $p < .05$; Kropp & Hart, 2000). Kropp and Hart also reported the following reliability statistics for the complete SARA: item heterogeneity ($N = 2,309$) with a mean inter-item correlation (MIC) of .15 and internal consistency with a Cronbach's α of .78.

The validity of SARA assessments have been more widely studied than many of the other measures included in the present review (see Tables 1–7). Five of the seven SARA articles in our review presented convergent validity data (i.e., correlations between SARA assessments and other relevant measures). When convergent validity of SARA assessments has been studied, results generally have revealed large, significant correlations (Cohen, 1988) in the expected directions with other IPV risk assessment instruments, such as the ODARA (Hilton, Harris, Rice, et al., 2008; Hilton et al., 2004) and the DVSI (Williams & Houghton, 2004). Kropp and Hart (2000) found that in Part 1 of the SARA, the items that pertain to general offending had moderate, statistically significant correlations with the Psychopathy Checklist: Screening Version (PCL:SV, Hart, Cox, & Hare, 1995; .45, $p < .001$), General Statistical Information on Recidivism Scale (GSIR, Nuffield, 1982; $-.40$, $p < .001$), and VRAG (Quinsey et al., 2006; .50, $p < .001$)—measures associated with general offending. When comparing the SARA total scores to the same general risk assessment instruments, the correlations are not as strong—PCL:SV (.43, $p < .001$), GSIR (-0.07, ns), and VRAG (.29, ns; Kropp & Hart, 2000). This is to be expected given that Part 2 and the SARA total score attend to IPV-specific risk variables.

Kropp and Hart (2000) reported criterion group validity in two ways. First, they compared mean total SARA scores for a group of inmates with a history of IPV

perpetration ($N = 638$) and a group of inmates without a history of IPV perpetration ($N = 372$), finding a significant difference between these groups: $M = 16.39$ ($SD = 6.86$) and $M = 7.51$ ($SD = 3.56$; $t = 27.04, p < .001$). Second, they compared the mean total SARA scores between probationers who recidivated ($N = 52$) and probationers who did not recidivate ($N = 50$), finding no significant difference between these groups: $M = 17.69$ ($SD = 5.60$) and $M = 15.68$ ($SD = 5.40$; $t = 1.85, p = .068$). Although the total score of the SARA did not significantly differ between the recidivist and non-recidivist groups, the Part 2 scores that pertain specifically to spousal assault risk factors were significantly different: $M = 7.64$ ($SD = 2.88$) and $M = 9.25$ ($SD = 2.61$; $t = 2.96, p = .004$). Kropp and Hart concluded that their results offer support for the primarily discretion-based SARA assessment reflecting the low, moderate, and high summary risk judgments. For instance, they found that of men who were recidivists, 60% ($n = 31$) had received a high summary risk rating, and just 8% ($n = 4$) had been classified as low risk.

The predictive validity of SARA assessments have been examined in several published studies. Nine AUCs were reported across all seven studies identified in our review. Using total scores of the SARA and various criteria pertaining to recidivism, the AUCs ranged from .52 to .65, exhibiting marginal predictive validity according to standards in the field (e.g., Hosmer & Lemeshow, 2000). Grann and Wedin (2002; discussed earlier) found that individuals scoring above the sample mean score of 20 on the SARA were almost 3 times more likely to recidivate than those scoring below 20 ($OR = 2.70$, 95% CI [0.95–7.66]).[14] Two studies examined the predictive validity of SARA summary risk ratings, yielding AUCs of .57 (Belfrage et al., 2011) and .70 (Kropp & Hart, 2000). Heckert and Gondolf (2004), who used a simulated version of the SARA in their study, reported a sensitivity of 43% and a rate of 27% false positives, whereas a study that scored the SARA using files reported a sensitivity of 82% and a specificity of 50% (Grann & Wedin, 2002).

In one of the most recent studies published to examine SARA assessments, Belfrage et al. (2011) compared SARA scores for recidivists and nonrecidivists of IPV. An important contribution of this study is that police officers completed the SARA assessments and then used them to implement risk management decisions. The correlation between the total SARA score and recidivism was small ($r = .18$, $p < .001$), and the ROC reported was good at .63; but as the authors asserted, in light of the risk management steps taken, it is reasonable to expect a smaller association. The ultimate purpose of any risk assessment is to reduce the risk of the anticipated adverse event. The Belfrage et al. (2011) study is a particularly important contribution not only to the validation of the SARA measure but also to the IPV field generally because they found that SARA's numerical total scores mediated the association between risk assessments and recidivism.

Taken together, the validation research on the SARA shows a versatile IPV risk assessment measure that can be used by a range of professionals, from clinicians to police officers, in a multitude of settings. All seven studies found in our literature search reported predictive validity results for SARA assessments. Overall, the results

indicate evidence of moderate predictive validity; however, given that the measure often has not been used in the manner intended (i.e., it was often coded from files alone), these results need to be interpreted cautiously. Considerably more research using prospective designs and coding the SARA based on comprehensive victim and perpetrator assessments will add essential information to the literature and may reveal more substantial predictive accuracy for SARA assessments. Unlike many other IPV risk assessment measures, the SARA authors also note that it has the potential to provide valuable information to inform risk management. Belfrage and colleagues (2011) published the first study that has examined how risk assessments can inform treatment planning, potentially leading to a reduction in recidivism, with promising results. The authors concluded that their findings indicate the "SARA is an effective tool not only for risk assessment and risk management, but also for violence prevention" (p. 66; see Table 5).

Brief Spousal Assault Form for the Evaluation of Risk (Kropp et al., 2005). The B-SAFER is a derivation of the SARA, developed especially to be used by law enforcement officials to assess risk of IPV (Kropp et al., 2005). Completion time is less than for the SARA, the measure does not require the assessment of a mental disorder, or a case history to complete—features that allow this succinct measure to aid in informing police of the risk of spousal assault while working within their realm of expertise and the practicalities of the contexts in which they must complete their assessments (Kropp et al., 2005). It should be noted, however, that there are two instances on the B-SAFER when the assessor is required to make a judgment pertaining to a psychiatric diagnosis; the presence of a mental health disorder and a substance use problem must be rated (Kropp et al., 2005), although an actual diagnosis is not needed for the item to be rated. The user qualifications include experience in individual risk assessment and knowledge in IPV (Kropp et al., 2005).

The development of the B-SAFER began with a systematic review of the IPV risk assessment literature (Kropp et al., 2005). A statistical analysis of the SARA was then conducted to address any possible overlap in items (Kropp et al., 2005). The SARA was pilot tested by law enforcement personnel in Sweden to determine if each of the items could be practically assessed in that setting, and the measure was further pilot tested by police in Sweden and Canada (Kropp et al., 2005). The B-SAFER was designed for use with adults who have a history of IPV, "regardless of gender or sexual orientation" (Kropp et al., 2005, p. 7). The measure addresses 10 risk factors for IPV, which were divided into two main sections. The five risk factors in the first section are intended to gather information about a perpetrator's history of IPV, whereas the second section addresses matters relating to the perpetrator's social and psychological functioning. The authors noted that the information needed to code this measure can be collected by interview or file; although they recommend that the clear preference would be to code the items from a combination of self-report, collateral information, personal observations, and a psychological evaluation (Kropp et al., 2005). The coding scheme for the B-SAFER consists of rating the presence or absence of risk

factors (Kropp et al., 2005). A "Y" means the risk factor is present, a "?" means that the rater is unsure if the item is present or the factor is possibly or partially present, and an "N" means the risk factor is absent (Kropp et al., 2005). Upon completion of the B-SAFER, the assessor introduces appropriate risk management strategies for the risk variables that are present (Kropp et al., 2005). In addition, reassessment of risk using the B-SAFER is recommended when there is a major change in the case or every six months to a year, but the time frame depends on the context and risk level of the individual being assessed (Kropp et al., 2005). For example, an individual who is at high risk for recidivism should be reassessed more often, whereas an individual who is institutionalized may not require reassessment as often (Kropp et al., 2005).

In our literature search, we found two published research articles on the B-SAFER that fit our inclusion criteria. Neither of the articles reported predictive validity data. The objective of the first study, by Belfrage and Strand (2008), was to test the inclusion of five victim vulnerability factors and to investigate whether victim vulnerability factors contributed to the police officers' risk assessments. In this case, the authors correlated the items with the SPJ of the police officers who rated the B-SAFER. The second objective of the study was to examine the prevalence of the victim vulnerability items and to determine if these factors could be coded in an acceptable way. The paper did not provide IRR, internal consistency, or predictive validity.

The mean age of the study group of 540 alleged male perpetrators of spousal assault was 38 years (range 13–76 years; Belfrage & Strand, 2008). The index offenses of which the aggressors were suspected included assaults (58%, $N = 312$), illegal threats (18%, $N = 95$), violations of a women's integrity (14%, $N = 76$), and other crimes such as attempted murder, molesting, and invasion of privacy (Belfrage & Strand, 2008). The victim vulnerability items were coded on the same 3-point scale as the B-SAFER ("Y" [*present*], "?" [*unsure*], and "N" [*not present*]), but they were only coded in the "present" situation (approximately the last four weeks), whereas the B-SAFER items were coded in both the "present" and "historically." The distribution of the victim vulnerability factors were as follows: inconsistent behavior/attitude ($M = 0.77$, $SD = 0.83$), extreme fear ($M = 0.82$, $SD = 0.80$), inadequate access to resources ($M = 0.47$, $SD = 0.70$), unsafe living situation ($M = 0.67$, $SD = 0.79$), and personal problems ($M = 0.61$, $SD = 0.80$; Belfrage & Strand, 2008). Both the B-SAFER items and victim vulnerability items correlated strongly with the degree of risk assessed by the police, leading the authors to conclude that the information on the measure is consistent with the variables police officers consider to be pertinent to IPV cases (Belfrage & Strand, 2008).

The second article we found on the B-SAFER, by Thijssen and de Ruiter (2011), had two main objectives: to (a) classify subtypes of spousal assaulters according to the typology developed by Holtzworth-Munroe and Stuart (1994) using four risk factors of the B-SAFER; and (b) evaluate the relationship between the subtypes of spousal assaulters and the recidivism rate. The IRR for the B-SAFER in this research was established using 12 cases after a one-day training workshop. The mean ICC was .57 (range: .21–.74). The items violent acts (ICC = .65), general criminality (ICC = .74), and substance use problems (ICC = .69) had good IRR. Only the item mental health

problems (ICC = .21) obtained a low reliability between raters; the authors attributed the poor reliability to insufficient information in the files.

Clearly, there is very little published research available on the B-SAFER to date, but the two available studies suggest that the measure has reasonable IRR and promising preliminary evidence for convergent validity (also see Au et al., 2008[15]).

Victim Risk Assessments—Structured and Unstructured

Danger Assessment Scale (J. C. Campbell, 1986; J. C. Campbell, Webster, & Glass, 2009). The original DA (J. C. Campbell, 1986) was developed in conference with women who had experienced IPV law enforcement personnel, shelter workers, and other experts on IPV to assess the risk of severe injury or homicide of a woman who is battered by a current or ex-intimate partner (J. C. Campbell, 1986). The DA originally was intended to be used as part of a nursing assessment in a clinical setting, although the author also asserts that it can be completed by the victim alone or with the aid of a health care worker or other advocate (J. C. Campbell, 1986). The measure was revised following an 11-city case-control study that compared victims of femicide and attempted femicide to abused controls from the same city (J. C. Campbell et al., 2009). The multivariate analyses from that study informed the revision of the second part of the instrument from 15 to 20 items as well as the development of an optional weighted scoring algorithm (J. C. Campbell et al., 2009; Hart & Watt, 2008).

In its present format, the first part of the DA assessment includes the presentation of a past year calendar, an important aid in assessing the frequency and severity of abuse that has occurred in the woman's relationship (Stuart & Campbell, 1989; also see http://www.dangerassessment.org/publications/aspx). The woman records the incidents of physical abuse on the calendar using a 5-point scale to indicate the severity of the abuse (e.g., 1 [*injuries that do not incur lasting pain*] and 5 [*use of a weapon*]). The second part of the DA is made up of 20 yes/no questions that yield a score between 0 and 20 (without the use of the weighted scoring) or between −3 and 37 (with the weighted scoring algorithm; J. C. Campbell et al., 2009; Hart & Watt, 2008). The unweighted scores are classified into four risk categories as follows: less than 7 represents a risk of variable danger, 8–13 represents a risk of increased danger, 14–17 indicates a risk of severe danger, and scores greater than 18 indicate extreme danger. A revised version of the instrument is also available (the Danger Assessment-Revised [DA-R]) that includes specific risk factors pertinent to female same-sex relationships (Glass, Perrin, et al., 2008).

Our systematic search revealed 11 published studies that reported validity or reliability information on the DA (see Table 6). Reliability of the DA was discussed in three articles that reported five α coefficients between .72 and .84 (McFarlane, Campbell, Sharps, & Watson, 2002; McFarlane, Parker, & Soeken, 1995; McFarlane et al., 1998). All three of these studies specifically addressed IPV during pregnancy. Two of the articles discussed the relationship between abuse during pregnancy and femicide (McFarlane et al., 2002; McFarlane et al., 1995). The other examined the severity

TABLE 6. Danger Assessment Scale (DA)

Study	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
McFarlane, et al. (1995) *USA*	Pregnant women *N* = 1203	*M* not reported; Most (70%) 20–29 yrs	Prospective cohort study	Physical abuse during pregnancy	16%	IRR: Not reported Internal consistency: α = .84	Index of Spouse Abuse (ISA): *r* = .79	Not reported
McFarlane, et al. (1998) *USA*	Pregnant victims of IPV; Ethnically stratified cohort *N* = 199	23 yrs	Prospective	Not relevant	Unknown	IRR: Not reported Internal consistency: α = .72	Physical items from the ISA: *r* = .75	Not reported
Goodman et al. (2000) *USA*	Female victims of IPV *N* = 49	Not reported	Prospective (3 months)	Repeat abuse	22%	Not reported	Not reported	A score of one standard deviation on the DA was related to a fourfold increase in likelihood of re-abuse (OR = 4.18) When compared with the CTS2: • Only the DA was significantly associated with reabuse

Weisz et al. (2000) *USA*	Female victims of IPV *N* = 177	31 yrs	Prospective (4 months)[a]	Victim report of severe violence at follow-up	Not reported	Not reported	Not reported	$R^2 = .09$ (*ns*)
McFarlane, et al. (2002) *USA*	Pregnant women *N* = 687	Not reported	Retrospective case control study	Attempted femicide, femicide, or abuse	N/A	IRR: Not reported Internal consistency: • Femicides: α = .80 • Attempted femicides: α = .74 • Abused controls: α = .76	N/A	Women abused during pregnancy had 3.08 greater likelihood of becoming an attempted or completed femicide victim
Heckert & Gondolf (2004) *USA*	Males admitted to batterer treatment groups (and their partners) *N* = 499	Not reported	Prospective (3 month follow-ups for 15 months)	Multiple outcomes: • repeat reassault • one time reassault • threatening reassault • controlling behaviour or verbal abuse	Repeat re-assaulters: 23% One-time re-assaulters: 12% Threatening reassault: 20% Controlling behavior or verbal abusers: 26% No abuse: 19%	Not reported	Not reported	AUC = .70

(continued)

TABLE 6. (*continued*)

Study	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Hilton et al. (2004) *Canada*	Male offenders $N = 589 +$ 100 cases for cross validation	38 yrs	Retrospective follow-up (4.8 yrs)	Any subsequent violent assault against an (ex-) wife or (ex-) common-law wife	30%	Not reported	ODARA: $r = .43$ ($p < .01$)	AUC = .59, 95% CI [±05]
Glass, Perrin, et al. (2008) *USA* DA-R	Female victims of IPV $N = 84$	37 yrs	Prospective (1 month follow-up)	Threatened or actual physical or sexual assault by partner or ex-partner	33%	Not reported	Not reported	Mean DA-R score for no violence group = 9.05 and violence group = 11.25 Each additional point on the DAS increased the odds of threatened or actual violence by: • 1.29 (unweighted)

Study	Sample	Age	Design	Outcome				Results
Glass, Laughon, et al. (2008) USA	Sample 1: Abused young adults N = 53 Sample 2: Young adult femicides N = 23 Sample 3: Adolescent femicides N = 5	S1: 19 yrs S2: 19 yrs S3: 16–17 yrs	Retrospective, case control	Femicide and abuse	Not reported	Not reported	Not reported	Femicide victims (5.4) had significantly higher total score on DA compared to abuse victims (3.8), $p < .05$.
Hilton et al. (2008) Canada	Sample 1: Male offenders with a history of IPV N = 303 Sample 2: Male offenders with a history of IPV N = 346	S1: 36 yrs S2: 35 yrs	Retrospective follow-up (5.0 yrs)	Any subsequent violent assault against an (ex) wife or (ex) common-law wife	S1 = 49% S2 = 41%	Not reported	ODARA: $r = .43*$ VRAG: $r = .21*$ PCL-R: $r = .36*$ DVRAG: $r = .46*$ DVSI: $r = .36*$ SARA: $r = .61*$ $*p < .001$	Sample 1: • Wife assault: $r = .12$ Sample 2: • Not reported w/ DA Combined Sample IPV recidivism: AUC = $.56, p < .05$

(continued)

TABLE 6. (*continued*)

Study	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Campbell et al. (2009) *USA*	Attempted femicides *N* = 194 Femicide cases: *N* = 310 Abused controls: *N* = 324	≥ 18 yrs	Retrospective, case control	Attempted femicide, femicide, or abuse (physically assaulted or threatened with a weapon by a current or former intimate partner in the past 2 yrs.)	Not reported	Not reported	Not reported	Comparison of attempted femicide group to control group DA scores: • AUC = .92, *p* < .001 • Sensitivity .55 – .99

Note. IRR = inter-rater reliability; AUC = area under the curve; CTS = Conflict Tactics Scale (Straus et al., 1996); IPV = Intimate partner violence; DVRAG = Domestic Violence Risk Appraisal Guide; DVSI = Domestic Violence Screening Instrument; ODARA = Ontario Domestic Assault Risk Assessment; PCL-R = Psychopathy Checklist – Revised; SARA = Spousal Assault Risk Assessment; VRAG = Violence Risk Appraisal Guide.
ªVictims were requested to report likelihood of reabuse in the next year but follow-up occurred at 4 months.

of IPV experienced by pregnant women and their perpetrators' access to firearms (McFarlane et al., 1998). Four of the studies examined convergent validity of the DA with other relevant measures of IPV (see Table 6). These studies reported a range of correlations between $r = .21$ ($p < .001$; VRAG) and $r = .79$ (Index of Spouse Abuse-Physical Scale [ISA-P]; McFarlane et al., 1995).

Nine of the 11 studies that included the DA examined predictive/postdictive validity (see Table 6). A retrospective study, using "attempted femicide or femicide" as the outcome variable resulted in an AUC of .92 ($p < .001$) when comparing DA scores for victims of attempted femicide to those of an abused control group (J. C. Campbell et al., 2009). Goodman, Dutton, and Bennett (2000) used the DA to examine its ability to predict IPV reabuse in a prospective study of arrested batterers. Over the 3-month follow-up period, 22% of the 49 women reporting to an intake center were either threatened or revictimized. They found that one standard deviation change in the women's DA scores corresponded with an approximate fourfold (4.18) greater likelihood of reabuse ($M = 7.24$, $SD = 2.56$; Goodman et al., 2000).

Two articles that examined the measure's predictive validity scored the DA by file, and the authors noted the serious limitations in scoring the measure in the absence of victim input. Both of these studies evaluated the ability of the DA to predict IPV reassault and reported fair AUCs of .59 (95% CI $\pm$ 0.05; Hilton et al., 2004) and .56 ($p < .05$; Hilton, Harris, Rice, et al., 2008). In another study using a modified approach to code the DA, Heckert and Gondolf (2004) compared what they referred to as "simulated versions" of the DA, KSID, and SARA as well as women's perceptions of reabuse. In a sample of 499 women who were abused, they found that of the four approaches, the DA most accurately predicted repeat reabuse over 15 months (AUC = .70; sensitivity = 66%, false positive rate = 33%; Heckert & Gondolf, 2004). In yet another study using a modified version of the measure (12 items from the DA), Weisz, Tolman, and Saunders (2000) studied 177 female victims of IPV. Using victim reports of severe reassault at the four-month follow-up, the measure yielded an R^2 of .09 ($F = 1.39$, ns). As noted previously, Glass, Perrin, and colleagues (2008) revised the DA to be used in female same-sex relationships. With each additional point on the DA-R, the odds that the individual was threatened or the victim of physical violence during follow-up increased by 1.29 for the unweighted score and 1.21 for the weighted score (Glass, Perrin, et al., 2008).

The DA has been an important addition to the IPV risk assessment field. It provides victims of IPV with a structured tool to inform decision making based on the level of risk their situation may present. Research suggests that the DA may in fact increase a victim's understanding of her level of danger. Specifically, when a correlation was run between women reporting their perceptions of danger before filling out the DA and their actual scores on the DA, the result is a weak insignificant result ($r = .24$, $p \leq .1$), suggesting that taking part in a DA assessment changes women's views of their safety (Stuart & Campbell, 1989). Further research to demonstrate whether the dynamic change results in an *increase in the DA's predictive accuracy* is needed (i.e., incremental validity).

Despite the promise of the DA and the fact that this measure has the largest body of literature behind it, there are limitations in the research that inhibit a clear picture of the psychometric properties of the measure, thus far. In many of the studies, it is not entirely clear which version of the instrument is being tested. Similarly, the calendar portion of the DA often is not mentioned in the methods and procedures sections of the publications, making it unclear if this important aspect of the measure was used in the research.[16] Researchers and clinicians alike should use the calendar when possible and evaluate it further in research because preliminary evidence suggests that this step may well be essential to supporting women's recollections of repeated events over time and informing the risk assessment process (Stuart & Campbell, 1989).

Perhaps the most important limitation to note about the DA research is that despite the fact that the DA was developed specifically to assess for femicide risk, the bulk of the published literature has examined the DA's use as an assessment of IPV reabuse. Only three (J. C. Campbell et al., 2009; Glass, Laughon, Rutto, Bevacqua, & Campbell, 2008; McFarlane et al., 2002) of the 11 articles used femicide as an outcome variable, and they appear to have all drawn their data from the same retrospective parent study (Campbell et al., 2003). Six (55%) of the DA studies were prospective, whereas the remainder relied on data collected retrospectively. In addition, five of the studies did not use a complete version of the DA (Campbell, 1986, contained 15 items; J. C. Campbell et al., 2009, contains 20 items), one study relied on a simulated version of the DA (Heckert & Gondolf, 2004), and two studies coded the DA from file alone (Hilton et al., 2004; Hilton, Harris, Rice, et al., 2008). In sum, the DA appears to have considerable promise and fills an important void in the literature given that lethal violence is of primary concern, but we look forward to further empirical validation work. In particular, progress with the measure has perhaps been hampered as a result of the lack of a cohesive manual and the use of various versions of the measure across the available published studies.

Victim Appraisals. In the IPV risk assessment literature, there has been considerable discussion regarding the ability of victims to predict the likelihood that their partner will commit violence against them again in the future in the absence of employing a measure to structure that determination (Weisz et al., 2000). Hart (1994) suggested that the victim would be the best person to accurately assess future violence because they know their partner better than anyone (also see M. A. Dutton, 1996). Before the accuracy of victim prediction had ever been tested, there had been a small number of surveys that found that victims might be able to sense or predict when violence would recur (e.g., Follingstad, Laughlin, Polek, Rutledge, & Hause, 1991; Walker, 1984). Experts, however, cautioned that in some situations, victims may underestimate their risk based on life circumstances or as a reaction to trauma (J. C. Campbell, 1995; M. A. Dutton & Dionne, 1991).

Following debates in the 80s and 90s, researchers have begun to evaluate the validity of victim reports as a means of informing IPV risk assessments. In total, our search yielded seven articles that provided information about the validity of victim appraisals (see Table 7). In each of these studies, the victim's prediction was assessed

TABLE 7. Victim Appraisals

Study	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Weisz et al. (2000) USA	Female victims of IPV $N = 177$	31 yrs	Prospective (4 months)[a]	Victim report of severe violence at follow-up	Not reported	Not reported	Not reported	Bivariate analysis: • $X^2 (9, N=177) = 34.30, p = .000$ When victims' predictions were added: • R^2 increased significantly from .15 to .25 Regression analysis including indicators from DA w/ survivors' predictions: • R^2 increased from .09 to .22
Cattaneo & Goodman (2003) USA	Female victims of IPV Baseline: $N = 169$ Follow-up: $N = 96$	No mean reported; 18+ yrs for inclusion	Prospective (3 month follow-up)	Repeat abuse within 3 months	11.5% of participants available for follow-up reported the occurrence of repeat abuse	Not reported	Not reported	Victim's predictions were found to be significantly related to re-abuse • Sensitivity = 74% • Specificity = 58% • PPV = 40% • Cases accurately classified = 66%

TABLE 7. (continued)

Study	Sample	Age (M)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Heckert & Gondolf (2004) U.S.A	Male batterers admitted to batterer treatment groups & their partners N = 499 male batterers N = 348 female partners	Not given	Prospective (3 month follow-ups for a period of 15 months)	Multiple outcomes: • Repeat reassault • One time reassault • Threatening reassault • Controlling behaviour; or verbal abuse	Repeat re-assaulters: 23% One-time re-assaulters: 12% Threatening reassault: 20% Controlling behavior or verbal abusers: 26% No abuse: 19%	Not reported	Violence is likely: • AUC = .64 Perceptions of safety: • AUC = .63	Re-assault: AUC = .83 When women's charac-teristics but not their perceptions are taken into account: • AUC = .79 • Sensitivity = 58%

| Cattaneo et al. (2007) USA | Female victims of IPV
$N = 246$ | 33 yrs | Prospective (18 month follow-up) | Repeat abuse within 1 year | 26% of participants experienced re-abuse | Not reported | Not reported | 66% of sample assessed their risk accurately
• Victims were equally able to predict future abuse and no future abuse
• Victims of abuse were more likely to correctly predict abuse than not ($p = .01$) |
| Bell et al. (2008) USA | Female victims of IPV
$N = 244$ | 33 yrs | Prospective (18 month follow-up) | Repeat psychological abuse within 1 year | 52% of the sample reported being psychologically abused during the 18 month follow-up | Not reported | Not reported | 62% of participants assessed their risk accurately
Multivariate analysis found the following significantly improved the fit of the model:
• Diagnosis of PTSD
• Recent violence
• Psychological abuse
• Past stalking |

(continued)

TABLE 7. (*continued*)

Study	Sample	Age (*M*)	Design (Duration)	Criterion Variable	Recidivism	Reliability	Concurrent/ Convergent Validity	Predictive/ Postdictive Validity
Connor-Smith et al. (2011) *USA*	Female victims of IPV *N* = 728	31 yrs	Cross-sectional	Compared 2 groups with equivalent ODARA sores 1) Examined the likelihood that the presence of risk factors was associated with victims' ratings of high or low risk, respectively	Not reported	IRR: Qualitative explanations for victim perceived risk of future aggression: • High-risk explanations: % agreement *M* = 96% (range = 87%–99%) • Kappas ranged from .48 to .96 (*M* = .78)	Modest agreement between victim rating and ODARA dichotomized score: *K*= .34, 67% agreement: • 38.7% agreed low-risk; 28.3% agreed high-risk; 12.6% victims rated high risk, but ODARA rated low-risk; 20.3% victims rated low-risk, but ODARA rated high-risk.	Not reported

2) Qualitative explanations from victims for ratings of high and low risk

3) Examined agreement between victims' ratings of high/low risk with dichotomized ODARA risk assessment: For disagreements, compared 2 groups with equivalent ODARA scores

- Low-risk explanations: % agreement ranged from 91%–99% (M = 96%)
- Kappas ranged .70 to .97 (M = .80)
Internal consistency: Not reported

- Two groups with equivalent ODARA sores compared
- Victims who perceived higher risk reported higher levels of ODARA dynamic risk factors

Note. IRR = inter-rater reliability; AUC= area under the curve; PPV = positive predictive value; IPV = Intimate partner violence; DA = Danger Assessment Scale; ODARA = Ontario Domestic Assault Risk Assessment.
[a]Victims were requested to report likelihood of reabuse in the next year but follow-up occurred at 4 months.

by a question(s) asking them to rate the likelihood that their partner would engage in a specified outcome behavior (e.g., physical violence, psychological abuse) during a specified follow-up period. All of these studies were carried out in the United States, and the sample sizes varied from $N = 96$ (Cattaneo & Goodman, 2003) to $N = 728$ participants (Connor-Smith, Henning, Moore, & Holdford, 2011).

Two articles used data from the same sample of women to assess the accuracy of victim reports—one for physical reabuse and the other for psychological reabuse over one-year follow-ups (Bell, Cattaneo, Goodman, & Dutton, 2008; Cattaneo, Bell, Goodman, & Dutton, 2007). The generalizability of both studies is limited to the extent the sample reflected a subset of predominantly impoverished African American women. Fisher's test yielded significant results for both the physical ($p < .01$) and psychological reabuse ($p < .001$) studies, suggesting that victims were more likely to be correct than incorrect (i.e., for dichotomous outcomes of true positive/ true negative vs. false positive/false negative). However, McNemar's chi-square tests were both insignificant with $p = .23$ and $p = .53$ for the physical and psychological reabuse groups, respectively (examining true positive, true negative, false positive, and false negative[17]). Bell et al. (2008) concluded that 62% of victims were "correct" in their assessment of risk of future psychological abuse. Cattaneo et al. (2007) reported that among their sample of women ($N = 246$) seeking help for IPV, victims were more likely to be right than to be wrong when asked to assess their risk of physical reabuse. Recidivism occurred within 26% of the participants' relationships during an 18-month follow-up period. Of the 246 participants, 134 were classified as true negatives, and 64 were classified as true positives. Overall, the authors concluded that there was no evidence of systematic bias in women's appraisals, meaning the women were no more likely to overestimate or underestimate their partners' risk of future IPV. Importantly, however, the results did provide insights into the circumstances or victim profile that might be indicative of a need for caution in making use of victim assessments. The results suggested that symptoms of substance use and posttraumatic stress disorder (PTSD), prior experiences of stalking or severe violence, and the temporal distance of the most recent incident of IPV are important areas for further inquiry in victim appraisal research. For instance, women with a history of substance use were more likely to underestimate their risk (false negatives).

Connor-Smith et al. (2011) compared victim assessments of risk to scores from a simulated version of the ODARA ($N = 728$); the victims' perceived risk appraisals were measured by asking if the women thought that their perpetrators would be violent with them in the next year. A "Yes" was considered indicative of a high summary risk rating, whereas a "No" was considered indicative of a low summary risk rating. There was modest agreement between the two risk assessment methods with 67.0% of participants' risk ratings in agreement with the simulated ODARA score. A high victim report risk estimate was reported by 12.6% of participants who had a low-risk ODARA score. Conversely, 20.3% of participants rated their risk of reabuse as low, whereas the ODARA score suggested they were at high risk of reassault. Victims who

reported a higher risk estimate than had been indicated on the simulated ODARA tended to report a greater incidence of dynamic risk factors than the low-risk group whose results indicated agreement between the two risk assessment approaches. The discrepancy between victims' reports and simulated ODARA scores could be because the ODARA was coded predominantly from criminal records where dynamic risk factors might not be so readily available (Connor-Smith et al., 2011). Another serious limitation to the study is that participants were not followed up and the actual incidence of reabuse is unknown.

Cattaneo and Goodman's study (2003) examined predictors of reabuse to include scores on the physical abuse, sexual abuse, and injury subscales of the CTS; employment history and socioeconomic status (SES) of abuser; along with the victims' assessment of reabuse over three months. Victims' predictions were found to be significantly related to reabuse, yielding a sensitivity of 74%, specificity of 58%, positive predictive value (PPV) of 40%, and the percentage of cases that accurately classified the risk level was 66%.

Two studies examined victims' appraisals against other methods of IPV risk assessment (Heckert & Gondolf, 2004; Weisz et al., 2000). Weisz et al. (2000) added victims' predictions to a multiple regression model of risk factors and a separate analysis with 12 items from the DA. In both cases, the addition of the victim appraisal item (which had the strongest bivariate relationship, $r = .42$) yielded a stronger more significant model for risk assessment (Weisz et al., 2000). Specifically, the multiple regression model initially yielded an R^2 of .15 ($F = 2.06, p < .05$), but with the addition of the victim report rating, the R^2 increased to .25 ($F = 3.56, p < .001$). The R^2 increased from .09 ($F = 1.39, ns$) to .22 ($F = 3.52, p < .001$) for the model with the 12-item DA (Weisz et al., 2000).

Heckert and Gondolf's (2004) study used a model with multiple outcomes to include repeat reassault, one-time reassault, threatening and controlling behavior, and verbal abuse to compare the predictive validity of victim assessments across various criterion variables. They reported moderate-to-strong evidence for the predictive ability of the following categories: men's characteristics and reports (AUC = .75), men's and women's characteristics and reports (AUC = .79), and men's and women's characteristics and reports combined with victim prediction (AUC = .83). Second, they compared victim report to the predictive validity of the KSID, SARA, and DA. The women's perception that violence was likely in the next three months (AUC = .64) was more predictive of reassault than the KSID (AUC = .57), the same as the SARA (AUC = .64), and less than the DA (AUC = .70), although the authors did not report specific analyses to compare the effect sizes. When the DA and the SARA were combined with the women's perceptions of violence in the next three months, both AUCs increased (to .73 and .69, respectively); suggesting that women's appraisals may add unique and valuable information, although incremental validity was not tested to determine if these increases were significant.

There is considerable variability in the procedures researchers have used to collect victim report data, making it difficult to directly compare results across this

small body of research. For instance, some studies have asked women to assess their risk following the completion of other measures, whereas other studies have simply asked single, open-ended questions. In addition, the context for follow-ups varies (e.g., duration, in person vs. phone calls, the nature of risk that constitutes reabuse). Even within study methodological rigor is poor. As mentioned by Weisz et al. (2000), in that study, the victim prediction was obtained at the end of an in-depth interview about the victim's IPV victimization history, thereby potentially contaminating the victim appraisal with a whole host of other variables. A particularly poor design feature, evident in several of the studies, is the apparent mismatch between the time frame in which the victims were asked to determine their partner's risk to perpetrate future abuse and the actual follow-up period (also see Bowen, 2011).

In summary, women's appraisals appear to contribute valid and potentially unique information to IPV risk assessments (see Table 7). Victim appraisals of the risk of future IPV show some evidence of predictive accuracy, even exhibiting greater predictive validity than some risk assessment instruments (Heckert & Gondolf, 2004) and adding significantly to regression models (Weisz et al., 2000). The research reflects various different methods of assessing victim reports, making it difficult to compare between studies. As Hanson et al. (2007) reminded us, it is unknown what each victim is taking into account when assessing her personal risk. Neither the individual risk variables that victims considered nor the method of combining risk factors into an overall level of risk is known, making it difficult to study victim appraisals; although research is starting to move in this direction (e.g., Connor-Smith et al., 2011). Finally, there is insufficient evidence to confirm if there are subgroups of women (e.g., with PTSD, substance misuse) for whom victim appraisals might consistently overestimate or underestimate risk of future IPV, but this will have important clinical implications.

Other Methods and Measures (Pilot Studies)

Our review also revealed a small number of newly developed measures and pilot studies.

Danger Assessment Scale Brief Risk Assessment for the Emergency Department (Snider, Webster, O'Sullivan, & Campbell, 2009). Snider and colleagues (2009) developed a brief risk assessment measure derived from the DA (J. C. Campbell, 1995) for use in emergency departments. The authors used data from a larger study called the Risk Assessment Validation Evaluation (RAVE; Roehl et al., 2005) to develop this new measure. The sample consisted of 666 recent female victims of IPV who were recruited from locations such as family courts, hospitals, shelters, and sheriff's departments in New York City and Los Angeles. The mean age of the women who completed both the baseline and follow-up interview was 31.9 years (95% CI [31.00–32.70]), and 39.4% were born outside of the United States. The sample was predominantly Hispanic (56.0%) and Black/African American (26.8%), limiting the generalizability of the results. The women were interviewed between 2002 and 2004

and 400 completed a follow-up interview after nine months, on average. The outcome measure was an assault by a partner or an ex-partner that inflicted severe injury or was potentially fatal. The outcome measure was based on the CTS2 (Straus & Douglas, 2004) as well as some questions developed for the RAVE study.

To identify which of the DA items were most predictive, a logistic regression analysis was used with manual backward stepwise selection of the items. The analysis resulted in the inclusion of five items from the DA in the new instrument. These five items (theoretical total score ranged from 0 to 5) had a predictive validity of AUC = .79 (95% CI [0.73–0.85]). When three out of five responses were "Yes," the sensitivity was 83% (95% CI [70.6%–91.4%]), specificity was 56% (95% CI [50.8%–61.8%]), and the PPV was 25%. Cross-validation research is required to determine if the measure yields further promising results in other samples.

Structuring Clinical Judgment. Sampling from four U.S. cities, Gondolf and Wernik (2009) intended to evaluate the validity of clinicians' assessments to predict the risk of reassault in a group of men participating in batterers' treatment programs ($N = 854$). The clinicians were requested to rate 10 items relevant to the men's behaviors in the programs and how they were performing. The 10 items fell into three constructs: (a) treatment adherence factors (i.e., attendance, using techniques, help seeking, active engagement), (b) problem behaviors (i.e., nonviolence, sobriety), and (c) changes in psychological mindfulness (i.e., acceptance, process consciousness, self-disclosure, use of sensitive language). The items were scored on a 5-point Likert-type scale ranging from 1 (*very little present*) to 5 (*extremely present*). Recidivism included reassault and severe reassault reported by the men's female partners during follow-up interviews (at six months posttreatment and 15 months postintake follow-up) using the CTS2 (Straus et al., 1996). Severe reassault was defined as severe tactics on the CTS2 (e.g., hit with a fist, bit, kick; hit with something or attempted to hit with something; choked or burned; threatened with a knife or gun; used a knife or gun; forced sex against will). Internal consistency was very high ($\alpha = .97$). In general, the logistic regression and ROC analyses suggested modest predictive validity for the clinical ratings. The programs with a longer duration had higher effect sizes in predicting severe reassault as follows: 9-month program: AUC = .74 (*SE* = .08, 95% CI [0.59–0.90], $n = 134$); 5.5-month program: AUC = .65 (*SE* = 0.06, 95% CI [0.55–0.76], $n = 113$). The authors concluded, "the sums of clinicians ratings are significant but weak predictors of especially severe reassaults" (Gondolf & Wernik, 2009, p. 1792). Given that the rating sum was a significant predictor of "severe reassault" only for completers of the programs ($R^2 = .03$, $\chi^2 = 4.14$, $p < .05$, $n = 34$), the authors postulated that motivation may be a key factor.

Severe Intimate Violence Partner Risk Prediction Scale (SIVIPAS; Echeburúa, Fernández-Montalvo, de Corral, López-Goñi, 2009). The SIVIPAS (Echeburúa et al., 2009) is a new actuarial scale developed in the Basque Country, Spain. Although it is presented as a prediction scale, the items of the scale have the objective

to discriminate between severe violence and less severe violence.[18] A strength of the calibration study was the size of the sample (N = 1,081 male batterers).

The SIVIPAS consists of 20 dichotomous items divided into five subscales: (a) personal data, with only one item (male batterer or victim is an immigrant); (b) couple relationship status, with two items (recently separated, recent harassment, or violation of restraining orders); (c) type of violence, consists of seven items (physical violence that causes injuries, physical violence in presence of children, severe threats to kill victim in past month, threat with weapons); (d) male batterer's profile, also with seven items (intense jealousy, history of violent behavior with previous partner, abuse of alcohol/drugs, justification of violent behavior due to aggressor's state because of alcohol, drugs, or stress); (e) victim's vulnerability, three more items that completes the 20-item scale (victim's perception of danger in past month, attempts to drop charges, or going back on decision to leave). After scoring the variables, the 20 items are summed up to create a total score. The interpretation of the total score is through cutoff scores into three levels of IPV severity: low risk (0–4 points), moderate risk (5–9 points), and high risk (10–20 points).

The 20 items, selected from a pool of 58, were able to distinguish between the severe violence group and the less severe violence group. The differences between the two groups was determined by univariate chi-square comparisons (Echeburúa, Fernández-Montalvo, & Corral, 2008) and not through multivariate analysis, which would have taken into account the associations between the predictor variables.

The development sample consisted of 269 (24.88%) aggressors with a less severe violent index offense and 812 (75.12%) aggressors who committed an index offense with severe violence. Data on previous offenses was not mentioned. In the severe violence group, the immigrants were more represented, especially Latin Americans (severe violence: 16.7%, less severe violence: 13.8%) and Africans (severe violence: 10.7%, less severe violence: 6.3%). Also, the victims of severe violence were more from Latin America (19.4%) and Africa (4.5%), compared to the victims of less severe violence (24.5% and 3.4%, respectively; Echeburúa et al., 2008, p. 364). The mean age of the aggressors was 37.3 years (SD = 10.4) for the severe violence group and 38.2 years (SD = 11.2) for the less severe violence group. The mean age of the victims of severe violence was 34.5 years (SD = 9.6) and for the second group, M = 35.1 years (SD = 10.8; Echeburúa et al., 2008). Members of the Basque police collected the data for the variables when the victims filed a complaint. The police assigned the aggressors to a group based on interviews and the particular circumstances surrounding their crime.

The internal consistency of the scale measured with all participants was α = .71 within the range of acceptable values (DeVellis, 2012). The internal consistency calculated for the subsamples of severe aggressors (α = .69) and the less severe aggressors (α = .66) were both in the minimally acceptable range (DeVellis, 2012). Internal consistency of the subscales was not mentioned. Regarding predictive validity, the authors reported a correct overall classification of 73.1%, with a sensitivity of 47.9% (the relative correct classification regarding the total of severe cases) and a specificity

of 81.4% (the relative correct classification of nonsevere cases regarding the total of nonsevere cases).

The SIVIPAS is a new actuarial instrument to discriminate severe violence from less severe violence; in this sense, it is not a risk assessment instrument to predict the likelihood of the occurrence of IPV recidivism. The instrument was developed to assist "police, social workers, forensic psychologists and judges in their decision-making process" (Echeburúa et al., 2009, p. 929). With only the development study and without a cross-validation study, the instrument remains in the pilot phase and would not be suited for application in daily practice. Moreover, the study did not report IRR, and the authors mentioned as one of the limitations that although the investigators attempted to train the assessors, they still may have used different criteria when assigning subjects to groups. Regarding the inclusion of the item "batterer or victim is immigrant," there could be a possible bias, immigrants are less protected and more vulnerable, and the reported results could be an artifact of the situation of (illegal) immigrants in Spain. If the violence is not serious (illegal), immigrants will probably be less likely to report the offense to the police. This could explain the somewhat lower presence of immigrants in the less severe violence group.

Los Angeles Sheriff's Department Screening Measure (Berk, He, & Sorenson, 2005). When the Los Angeles Sheriff's Department was searching for a tool for forecasting IPV, they were unable to find one that suited their unique needs. They wanted a short screening instrument that was valid for use among the diverse ethnic populations in Los Angeles County. In response, they initiated the development of a screening measure to be used by police who are encountering IPV cases in the field. The development sample consisted of 1,500 representative households to which the sheriff's deputies had been dispatched for incidents that involved domestic violence. Deputies initiated data collection with a 30-item list of questions selected because of their perceived importance in the prediction of past IPV incidents. For analysis, data-mining techniques and classification and regression trees (CART; Breiman, Friedman, Olshen, & Stone, 1984) were used; consequently, the screening instrument had the structure of a classification tree. Using a cost ratio of 5:1 for false negatives (incorrectly forecasting no future calls) to false positives (incorrectly forecasting future calls), four of the original 30 items predicted future calls to the sheriff's office 60% of the time and accurately forecasted the absence of domestic violence 50% of the time. These four predictors were (a) whether the victim reports that there have been more than three police calls to the household before, (b) whether the perpetrator is reported to damage household property when angry, (c) whether the perpetrator is reported to be unemployed, and finally, (d) whether the perpetrator is reported to have threatened the life of the victim or someone in the victim's family in the past. When the cost ratio was brought to a 10:1 ratio of false negatives to false positives, the accurate forecasting of domestic violence calls was 50% of the time, and the accurate forecasting of the absence of domestic violence calls was nearly 70% of the time (Berk et al., 2005).

DISCUSSION

This systematic review examined 19 different risk assessment measures[19] and reported on data from 39 separate publications. Most of the measures were developed with the intention of assisting diverse professional groups in identifying the risk of (male) perpetrated IPV (against women); however, we have also reported findings from general risk assessment measures[18] that have relevance to IPV risk assessments and have been studied with relevant populations (LSI/LSI-R, PCL-R, VRAG).

In one of the earliest published reviews in this field, D. G. Dutton and Kropp (2000) remarked that the "science and practice of spousal assault risk assessment is still in its infancy" (p. 178). We feel compelled to reiterate that firm conclusions regarding violence risk assessment for IPV remain somewhat elusive because of limitations of the research to date. Many of the advances evident in the larger violence risk assessment field have yet to be realized in the IPV field. In stark contrast to the publication of multiple meta-analyses in the general violence risk assessment field (e.g., Campbell, French, & Gendreau, 2009; Coid et al., 2009; Singh, Grann, & Fazel, 2011; Yang, Wong, & Coid, 2010), we found just one meta-analysis of the IPV risk assessment literature (Hanson et al., 2007). Similarly, compared to the exponential rate at which publications examining violence risk assessment for general violent offending and sexual offending are appearing in the literature, a fraction of studies are available that examine IPV. Guy (2008) reported on 113 violence risk assessment disseminations (although Guy's review included unpublished studies). Hanson and colleagues (2007) examined 88 studies of violent recidivism, and Singh, Grann, et al. (2011) examined 68 studies based on 25,980 participants. Of the 19 measures we examined, only 8 had been evaluated in three or more published studies and are no longer considered pilot instruments. The surge of research on violence risk assessments has simply not been matched in the IPV risk assessment field, which appears to be still very much in the early phases of development and dissemination. Most of the research on IPV risk assessment measures has been published quite recently (also see Hanson et al., 2007) and, as we elaborate in the following text, there are considerable methodological limitations hampering the clinical implications that can be drawn from the studies that have been completed to date.

Limitations of the Extant Literature and Implications for Research

Our review of the IPV risk assessment literature identified several limitations in the rigor of the research potentially limiting the generalizability of the present findings and the implications for clinical practice. These limitations also point to important avenues for future research. In addition to the review presented in this manuscript, a thorough summary of the sampling, procedures, and limitations of each study is provided in our online annotated bibliography.[20] Here we provide a brief discussion of some of the primary challenges hampering the field to date and provide concrete

recommendations for advancing the research and improving the body of work to better inform evidence-based practice.

Study Design. Much of the IPV risk assessment research is limited because of less than optimal research designs and study methods. The literature reviewed reveals a shortage of prospective, longitudinal studies. For example, of 11 publications on the DA, only 6 (55%) were prospective. There is also a need for research that directly compares multiple measures (Guo & Harstall, 2008; cf. Hilton et al., 2004). Our review found that just 9 of the 39 studies compared multiple risk assessment measures; this is problematic for several reasons. For instance, research demonstrates that prevalence estimates for IPV differ dramatically across studies. To demonstrate, in another systematic review, we recently found prevalence rates of physical IPV victimization ranging from 0% to 99%[21] (Desmarais et al., 2012a, 2012b). In addition, studies vary considerably in their definition of IPV, the method of completing the risk assessment measures, the sources of information used to complete the measures and to collect recidivism data, and the length of follow-up period used to calculate recidivism. The benefit of comparing multiple measures in the same study is that some of the methodological and contextual differences that make it difficult to compare across studies can be accounted for within a single project.

Quality of Risk Assessments. A particular limitation of the extant literature is that researchers often failed to administer the risk assessment measures in the way they were intended to be used (e.g., to solicit information from diverse sources, to use all items on a measure, to have the coding completed by the parties intended [e.g., police officers]). For instance, many studies of the SARA used file reviews alone in the absence of completing interviews with the perpetrator or obtaining any collateral information. In some studies, investigators relied on proxy versions of the measures, which means the data collected was not completed with the intention of coding the measure(s) at the outset. These limitations can lead to missing items and may underestimate the utility of the measures. In addition, most SPJ measures (e.g., SARA; Kropp et al., 1995, 2008) explicitly recommend that the risk assessment decisions are reflected in the summary risk estimates, but much of the predictive validity research has examined the actuarial use of the measure's total scores as opposed to reporting the accuracy of the risk judgment.

Quality of Outcome Criteria. A small number of studies have examined victim reports of reabuse as the outcome variable when evaluating predictive validity ($k = 7$; 18%). In our view, including victims and collaterals in collecting data on recidivism is likely particularly important in this field of research given that IPV occurs in the privacy of the home and in the context of relationships, which oftentimes are still intact. A persistent problem in the measurement of the criterion variable in the IPV risk assessment field is the reliance on domestic violence reoffending from criminal records and other official administrative databases (Hilton et al., 2010). This

method of collecting outcome data is problematic because it likely results in a vastly underestimated rate of recidivism. In addition, because criminal records do not identify victim–offender relationships, many studies using these records are restricted to using general violent recidivism as an outcome variable as opposed to IPV reabuse specifically, thereby seriously restricting the relevance of the findings for this body of research (e.g., Hilton et al., 2001). The limitations of the criterion data in the IPV risk assessment literature may have contributed to lower IRR for domestic violence recidivism. Missing data on the victim–offender relationship can also result in unwarranted conservative predictive accuracy statistics (Hilton & Harris, 2009). The outcome criterion often is also a poor fit with the intended use of the measure being studied, thereby limiting the results we can draw from the extant literature. An example is the DA which is intended to identify femicide, but eight of the 11 articles reporting on the DA have used abuse as an outcome rather than femicide.

The challenge of drawing conclusions based on the extant IPV risk assessment literature can best be illustrated with a specific example: Although both victim appraisals and the DA show promise in assessing risk, it is difficult to compare between them because both bodies of research reflect such a wide range of different study methods (e.g., different outcome variables, follow-up times). This is evident when looking at the AUCs reported for the DA, which varied from .56 to .92. The criterion variables differ between measures with the DA mainly assessing femicide and severe abuse, whereas the victim report studies used a range of outcome variables from psychological reabuse to severe violence. In light of the promise evident in having victims report on their risk of reassault, a next step would be to directly compare the incremental value achieved when women's appraisals are supported by the DA. Another avenue for further exploration is combining a victim report measure with one that focuses on the perpetrator's characteristics (e.g., SARA, ODARA) to see if the predictive validity is magnified.

A notable gap in the violence risk assessment field generally, as opposed to the IPV risk assessment field specifically, is limited research to date that demonstrates the use of risk assessment for decreasing violence risk and increasing community and victim safety, ultimately the objective of any risk assessment. Belfrage and colleagues (2011) took risk assessment research to the next level in their study of the SARA by evaluating the use of a risk assessment measure to inform risk management strategies. That same team has also advanced risk management and violence prevention in their work on the SARA and B-SAFER (Kropp et al., 2005; Kropp et al., 1995) by including risk factors that lead to scenarios and risk management planning.

A final note about criterion variables, experts in the field have also recommended that risk assessment research move beyond the very broad category of reabuse or recidivism used in most studies. Grann and Wedin (2002) cautioned that "the word 'risk' is not just the probability of a re-occurrence of an unwanted event; the nature and the severity of recidivism are just as important to consider" (p. 19). Indeed, the authors of the SARA and B-SAFER (Kropp et al., 2005) include scenario planning as part of risk assessments to encourage assessors to consider the nature, frequency, severity, and imminence of a violent event.

Samples and Generalizability. To date, very little research has been focused on subgroups who either perpetrate abuse and/or are victims of IPV. Of the 39 articles discussed in this review, only two included female perpetrators in the context of heterosexual relationships (Williams, 2012; Williams & Grant, 2006), whereas one included female perpetrators in the context of same-sex relationships (Glass, Perrin, et al., 2008). Another restriction posed by the current state of the IPV risk assessment field is that there are only two studies that included victims and/or perpetrators from same-sex couples as participants in the validation research, making it difficult for clinicians to know how to best assess risk among this population (Dutton et al., 2001; Glass, Perrin et al., 2008). Future research should include male victims, female perpetrators, diverse cultural and ethnic groups, and GLBT victims and perpetrators as well as including participants from diverse age groups (adolescents, older adults) and socioeconomic strata. Examining IPV risk assessment in diverse populations is important given considerable research evidence that abuse in intimate relationships is found across all sociodemographic contexts, heterosexual and homosexual couples (Glass, Perrin, et al., 2008), and is particularly prevalent among young, cohabitating, and dating couples (e.g., Desmarais et al., 2012a, 2012b).

Allegiance. Consistent with Guy (2008), we found that measures are most often tested within their own backyard. Studies were predominantly conducted in the United States and Canada (although as recalled, we restricted our search to publications in the English language) and often by coauthors of the measures. The SARA and DA are exceptions to this rule. Of seven cross-validation studies on the SARA, only two were coauthored by the manual developers. The DA has an impressive 11 publications (although three are from the same dataset/parent study), five of which were published by at least one study coauthor.

Novel Areas in Need of Research. In addition to limitations evident in the extant research, there are also aspects of IPV risk assessment that are ripe for future study simply because they are novel areas that have not been investigated sufficiently, or often at all. Few studies have examined the clinical overrides provided in some measures (see Girard & Wormith, 2004). There is some evidence that there may be some important key/critical items (e.g., on SARA) that point to a group of outliers who are at particularly high risk for IPV (e.g., violate custody orders). Grann and Wedin (2002) noted that individuals with a personality disorder and a past violation of a conditional release were three times more likely to recidivate than others. More research is needed to confirm whether or not this finding holds across other samples.

We found no studies that examined the self-reported risk of the perpetrator, although this might be a very worthy area of investigation. Skeem (2005), for instance, found that self-appraisals of violence risk might outperform professional risk assessments, suggesting this would be a potentially profitable avenue for further investigation. Finally, there is considerable debate in the risk assessment field presently regarding the extent to which training on measures such as those studied in

the present review can be implemented successfully into clinical practice (Murrie, Boccaccini, Johnson, & Janke, 2008) and to what extent we see strong evidence of "field reliability" (i.e., do the psychometric properties of the measure hold up in real-world clinical settings when the measures are put to use by actual clinicians). Future research should endeavor to sample clinicians using risk assessment tools in clinical practice and diverse institutional and community settings. Considerably more work is needed on intimate partner homicide (Eke et al., 2011) and stalking (e.g., Is it preferable to use an IPV measure or a measure of stalking, or some combination of such instruments for cases of stalking involving former intimate partners?). Finally, we would like to see continued examination of the variables that are associated with the accuracy of victim appraisals (e.g., Heckert & Gondolf, 2004; also see Bowen, 2011).

Limitations of Our Review

Our review did not include a methodological appraisal of the papers to determine the quality of the studies for inclusion. Although we completed a systematic search of the literature, we did not complete a meta-analysis and weight the indices of the associations between the measures and the outcomes because of the heterogeneity of the extant literature. In many cases, predictive validity analyses were not reported, and we identified several measures and approaches that have been evaluated in as few as one or two studies. We were inclusive given our intention was to provide a synthesis of all of the IPV risk assessment research that has been done to date, and our objective was to provide guidance for future research as well as clinical practice. As such, we used an inclusive search strategy to cover as much of the research as possible, although we limited ourselves to published studies. Our results also cannot be generalized to non-Westernized nations or to male victims or female perpetrators.

Implications for Clinical Practice

The Value and Necessity of Evidence Informed Intimate Partner Violence Risk Assessments. Risk assessment is considered the cornerstone of offender management in various legal (e.g., civil commitment, criminal sentencing) and clinical contexts (e.g., determining appropriateness for visit leaves and community access). Applying this body of knowledge to the IPV field is particularly important because the prognosis is generally poor for these offenders (Grann & Wedin, 2002). Research on treatment effectiveness with perpetrators of IPV indicates insufficient empirical evidence to demonstrate that one intervention is consistently superior to another or even that IPV treatment is particularly effective (Babcock, Green, & Robie, 2004; Grann & Wedin, 2002; see Gondolf, 1997; Hanson & Wallace-Capretta, 2004), suggesting that the risk of recidivism with this population is of considerable concern. Extensive research has also firmly established that the implications of IPV for victims, children, and society are extreme (as measured by psychological, physical, and economic outcomes; e.g., Reeves & O'Leary-Kelly, 2007; Walby, 2004) and not infrequently results

in intergenerational downward spirals for these families (e.g., Stith et al., 2000). Accurate IPV risk assessments have the capacity to reduce the prevalence, incidence, and severity of spouse abuse and support the allocation of scarce resources (e.g., close supervision of high-risk offenders, determine treatment needs). The benefits of using risk assessments also include facilitating the development and execution of appropriate safety plans for victims and providing a common language that can facilitate communication between professions and across organizations and settings. The focus of the assessment is to understand the variables and processes that may have led to abuse in the past and to understand what might influence future choices to engage in IPV or not. Ultimately, the objective is to improve the validity, reliability, transparency, and justifiability of risk determinations and interventions. The task that remains is to determine how best to complete an IPV risk assessment. Our review of the literature provides some guidance for making these critical determinations.

Reliability. IRR was tested for a minority of the measures (see Tables 1–7). Note for instance the DA has the largest number of studies of the measures we reviewed ($k = 11$), but it has not been evaluated for IRR, and we located just two studies providing support for the internal consistency of the measure (Table 6). Although it may be difficult to conceptualize IRR for the DA (given it is the victim's appraisal), we note that we did not locate a study reporting test–retest reliability. Of interest, however, Connor-Smith et al. (2011) reported 96% agreement between raters (Kappas range: .48–.96; $K = .78$) in their study of victim appraisals. The ODARA/DVRAG has only two studies and the DVSI-R has just one study reporting IRR, although each indicated strong interrater agreement (see Table 1). Limited research has provided evidence of the IRR of the PCL-R, LSI-R, and VRAG in the context of IPV; however, these generally have resulted in good IRR. In addition, there are large bodies of work outside the IPV area that support the reliability and IRR of the later three measures (Andrews et al., 2010a, 2012; Hare, 2003; Quinsey et al., 2006). The SARA is one of the most well-studied IPV risk assessment measures, and it has some of the strongest evidence to support its IRR. We found three studies evaluating the IRR of SARA total scores and each reported ICCs in the excellent range (also see Kropp & Gibas, 2010). Further examination of rater agreement for the summary risk estimates of the SARA is needed given that it is the intended approach for determining the offender's risk, but it has only been studied in one publication to date (Kropp & Hart, 2000).

Predictive Accuracy. Considering only those risk assessment measures/approaches that are no longer in the pilot phase and for which there are at least three published studies ($k = 8$; victim appraisal, ODARA/DVRAG, DA, SARA, LSI/LSI-R/LSI-OR, DVSI/DVSI-R, PCL-R, and VRAG; see Tables 1–7), our review suggests that the measures generally perform in the moderate to good range according to statistical standards (e.g., $r > .30$; Cohen, 1988; AUC $> .60$; Cicchetti & Sparrow, 1981; Rice & Harris, 2005; generally for a risk assessment perspective, see Otto & Douglas, 2010). For instance, the ODARA/DVRAG produced a single large effect size for predictive

validity but that was found in the construction sample (AUC = .77; Hilton et al., 2004); whereas the cross-validation studies, which reported data for these two measures ($k = 2$), either did not report predictive/postdictive validity indices or reported AUCs in the moderate range ($k = 5$; range = .64–.70). Similarly, the DVSI/DVSI-R resulted in small to moderate effect sizes depending on the outcome criterion considered. A consideration of the predictive accuracy of the DA is generally positive, demonstrating that attempted femicide victims receive higher scores than other victims of IPV, and the measure often results in substantial odds ratios. However, half of the DA studies have been retrospective or cross-sectional ($k = 15$ studies), and there are no prospective studies using attempted or completed femicide as the outcome criterion. Evidence for the SARA, an SPJ measure, would benefit from further research testing the instrument when administered in the intended manner (i.e., interviews + file reviews; collateral informants) and when analyses evaluate the predictive accuracy as recommended by the authors (i.e., summary risk judgments as opposed to actuarial/total scores). Presently, evidence for the validity of the SARA summary risk ratings comes from just two studies reporting AUCs of .57 and .70 (Belfrage et al., 2011; Kropp & Hart, 2000). The predictive/postdictive validity of the total actuarial score of the SARA is in the fair to good range based on the nine AUCs reported (range = .52–.65), whereas when the two correlations reported are examined, the literature only exhibits a small correlation between total SARA scores and recidivism ($r = .18$ and $r = .21, p < .001$). Victim appraisals have been studied in six publications considered here, five of which included some consideration of postdictive/predictive validity. Finally, the general risk assessment tools examined in our review appeared at first to be resulting in similar predictive accuracy. On closer scrutiny, in fact, there are only three published studies of the PCL-R and VRAG, and of those, just one study used IPV as part of the outcome criterion, thus seriously limiting the conclusions that can be drawn about the superiority of IPV-specific measures versus general risk assessment measures. We would also point out that when reflecting on the results of their metaregression analysis of 68 studies, Singh, Serper, Reinharth, and Fazel (2011) concluded that tools developed for use with particular offending populations performed better than general tools.

To our knowledge, there is no specific threshold to indicate that a measure has achieved a suitable level of predictive accuracy in terms of the number of studies that should have been published nor the average effect size that ought to have been achieved. If we rely on the general violence risk assessment field as a comparison, Guy (2008) reported that the mean weighted AUC of the SPJ approach using any antisocial behavior as the outcome of interest was .68 (nonweighted was nearly identical at .70). Similarly, using standard statistical criteria in the field as a comparison, AUCs of .68/.70 are generally considered to be in the moderate range. With these thresholds in mind, our review suggests that the predictive accuracy of IPV risk assessments is well within the range of the general violence risk assessment measures and may best be described as moderate and significantly better than chance.[22]

Predictive Superiority of Actuarial Versus Structured Professional Judgment. We generally found that the various approaches to identifying IPV risk levels had similar degrees of predictive accuracy (see Tables 1–7), although comparisons across the studies need to take into consideration methodological differences in the studies. There remains considerable controversy in the literature regarding the extent to which actuarial measures versus SPJ measures have superior predictive acumen. Contrary to claims in the literature (e.g., Quinsey et al., 2006) that professional discretion should play no role in violence risk assessments, some recent meta-analytic work has concluded that there is little evidence for the superiority of either actuarial or SPJ measures when both approaches are examined and compared directly (see Singh, Grann, et al., 2011). Specifically, Guy (2008) reported on 44 studies directly comparing actuarial and SPJ measures in the general violence risk assessment field (although she included unpublished findings). Yang et al. (2010) recently reviewed nine risk assessment measures and similarly asserted that the nine measures were highly comparable and as such,

> The selection of which tool to use in practice should depend on what other functions the tool can perform rather than on its efficacy in predicting violence. The moderate level of predictive accuracy of these tools suggests that they should not be used solely for some criminal justice decision making that requires a very high level of accuracy such as preventive detention. (p. 761)

Campbell et al. (2009) similarly concluded that "most measures are similar in their predictive power" (p. 580).

Conclusions

An increasing number of structured risk assessment measures are available to assist clinicians tasked with assessing the likelihood of IPV. Results from the research reviewed here generally support the use of IPV risk assessment measures, demonstrating that they significantly improve on chance accuracy. The body of empirical research supporting the clinical use of these measures is growing if one is tasked with assessing male-perpetrated IPV against women; however, considerably less is known about assessing the risk of female-perpetrated IPV and the relevance of these measures for use with GBLT partners. In addition, many measures are still in the piloting phase or the initial steps of cross-validation research, and as such are less suitable for implementation into clinical practice. Predictive validity across the measures and methods (actuarial, structure professional judgment, victim appraisals) generally ranges from fair to good. Considering the limited research, the methodological limitations of the available studies, and the relatively comparable evidence of predictive validity among instruments, we remain unconvinced that there is sufficient evidence to support the superiority of any one risk assessment tool for use in cases of IPV (also see Bowen, 2011; Guo & Harstall, 2008; Hanson et al., 2007).

In addition, the selection of a preferred risk assessment measure should reflect the circumstances of the case in combination with a consideration of the psychometric properties of the instrument.

When clinicians and administrators are faced with the challenge of determining which measure(s) to use to assess risk of IPV, they should carefully consider the *purpose* of the assessment (Heilbrun, 2009). For instance, is the objective of the evaluation simply to forecast risk or is it to identify treatment and supervision needs in which case an SPJ measure like the SARA is much more likely to offer necessary insights than an actuarial measure or a victim appraisal (also see Coid et al., 2009; Hanson et al., 2007). As Hanson and colleagues (2007) noted, none of these scales is intended to determine whether and to what extent a victim needs support,[23] and not all empirically supported risk factors are equally relevant to risk management. Assessors also should take into account the *context*, *setting*, and *resources* when evaluating which measure best suits their needs. For instance, some SPJ measures (e.g., SARA and PCL-R) may be more resource intensive than most actuarial measures making them inappropriate for certain circumstances (e.g., police responders; also see Coid et al., 2009).[24] In addition, many of the measures considered here require extensive professional training and *expertise* of the evaluator (e.g., PCL-R). Finally, consideration must be given to the characteristics of the *population* to be assessed (e.g., age, gender, ethnicity, SES) and the extent to which a measure has been cross-validated in similar samples is required (Heilbrun, 2009). When validation research and empirical evidence is limited, evaluators should be particularly cautious in the interpretation of results and should make the limitations clearly evident to stakeholders in their risk assessment documentation and communications. It is also important that measures are used as intended (e.g. administering measures and conducting interviews, reviewing criminal records and clinical files); therefore, if the setting and context does not lend itself to accessing the required information and sufficient time to complete each recommended component of an assessment, the measure may not yield accurate information. In particular, assessors want to be clear about the outcome of concern (verbal abuse, physical abuse, severe violence, stalking, femicide) and knowledgeable about relevant base rates (Heilbrun, 2009). Although lethal assault (which might reflect femicide, filicide, and/or familicide) is of greatest concern, the necessary evidentiary basis for recommending a measure to assess for risk of lethal IPV violence is highly limited (also see Bowen, 2011; Guo & Harstall, 2008; Hart & Watt, 2008).

The RNR model is considered the dominant model of "what works" for the assessment and treatment of offenders (see Andrews, 2012). Briefly, the core principles of the RNR model facilitate effective interventions through matching level of service to risk, assessing and targeting criminogenic needs, and tailoring interventions based on the individual strengths of the offender. Although the RNR model has been used predominantly in the general offending literature, we recommend it as a strong conceptual foundation to inform assessment and management of IPV offenders. Among the strengths of the model are the focus on individual needs, strengths, and the

supporting evidence base that shows significant reductions in recidivism when programs are in adherence with this model (Andrews, 2012). Recent IPV-specific research has also underscored potential drawbacks from applying disproportional interventions; that is, overintervening may encroach on the expression of an individual's own coping strategies, which may actually increase the likelihood of experiencing negative outcomes (e.g., recidivism; Belfrage et al., 2011) or decrease the likelihood that victims will seek services in the future. This finding is consistent with the general principles of the RNR model, which argue strongly against unnecessarily intensive services that are not only grossly wasteful of limited resources but also potentially harmful (see Dvoskin, Skeem, Novaco, & Douglas, 2012).

NOTES

1. Please note that although the authors acknowledge that men also frequently experience abuse in intimate relationships, and women are perpetrators at roughly equal rates as men, for the purposes of this article, we will refer to women as victims and men as perpetrators given that the risk assessment literature has focused almost exclusively on female victims of male-perpetrated violence in heterosexual relationships.

2. Although noted that Monahan (1981; see Webster & Hucker, 2007) made the key point that more information does not always yield more accurate assessments. A consideration of "a limited number of *relevant* and *valid* predictor items" is more important than an exhaustive examination that yields much irrelevant and ultimately confusing information (pp. 125–126). The point here is that collecting information about relevant and valid risk and protective factors from diverse sources is more likely to result in a valid and reliable assessment than obtaining information from one or a small number of sources.

3. Dr. Webster is a professor emeritus of psychology at Simon Fraser University/ psychiatry at University of Toronto and one of the world's most renowned experts in the topic of violence risk assessment. He has coauthored several of the most influential violence risk assessment measures including the HCR-20, a 20-item measure intended to assess risk for general offending and violence, which has been translated into 14 languages.

4. Note that the HCR-20 is intended to evaluate risk for general violence, not IPV.

5. This is akin to the approach used by insurance companies to determine payee rates.

6. Note that this resulted in us including one paper with a 2012 publication date: Williams (2012).

7. Receiver operating characteristic (ROC) analyses generate AUCs (see Mossman, 1994; Mossman & Somoza, 1991). Although they can be smaller than chance ($<.50$), AUCs generally range from .50 (chance prediction [no ability to distinguish individuals with vs. those without the outcome]) to 1.00 (perfect discrimination/accuracy). AUCs $>.90$ are considered outstanding discrimination, .80–.89 are excellent,

.70–.79 are acceptable, and .50 is equivalent to chance (i.e., the false positive rate is equal to the true positive rate; Hosmer & Lemeshow, 2000, p. 162; also see Cicchetti & Sparrow, 1981).

8. Based on Cohen's (1988) criteria for interpreting correlation sizes, correlations <.30 are considered small, moderate correlations range from .30 to .50, and correlations >.50 are considered large.

9. ICCs often are interpreted as <.40 = poor, .40–.59 = fair, .60–.74 = good, and .75–1.00 = excellent (Cicchetti & Sparrow, 1981).

10. Roehl et al. (2005) indicated the KSID has 10 items; Heckert & Gondolf (2004) described the KSID and indicated it has 11 items.

11. It is important to clarify that the PCL-R (Hare, 2003) is not a "risk assessment" measure but rather a measure of the construct of psychopathy.

12. These acts of violence ranged from first-degree murder to threatening. The data used in this study has been reported previously (Harris et al., 1993; Rice & Harris, 1995, 1997).

13. Coefficients of agreement range from −1 (total disagreement) to +1 (total agreement). Kappas are mathematically equivalent to ICCs, and thus can be characterized in a manner consistent with footnote no. 8 (Norman & Streiner, 2008).

14. Odds ratios are generally considered to not be biased by a restriction in the range of the predictor or the outcome variable. An *OR* ≥2.5 is generally taken to represent the lower limit of a strong association between predictor and outcome (Fleiss, Williams, & Dubro, 1986).

15. Note that this study conducted in Hong Kong does not fit our inclusion criteria: English publications from westernized nations.

16. Outside of the time range for this review, Stuart and Campbell (1989) found that the internal consistency of the DA rose from .57 to .64 when an item relying heavily on the use of the calendar was excluded. The authors noted that only 60% of the women had used the calendar when completing the assessment.

17. True negative = low-risk estimate, no reabuse on follow-up; false negative = low-risk estimate, reabuse on follow-up; false positive = high-risk estimate, no reabuse on follow-up; true positive = high-risk estimate, reabuse on follow-up.

18. Severe violence was defined as (a) a committed or attempted homicide, (b) the use of weapons or dangerous objects, (c) severe or repeated injuries that required both professional aid as well as hospitalization or continued medical assistance. Less severe violence was not specifically defined, and it is not clear if violence other than physical violence was included.

19. As noted earlier, the PCL-R is a measure of psychopathy, not a risk assessment tool, per se.

20. http://blog.springerpub.com/social-work/the-partner-abuse-state-of-knowledge-project-the-unprecedented-series-of-manuscripts-and-free-online-data-base

21. Male perpetration: 1.0%–61.6%, female perpetration: 2.4%–68.9%, male victimization: 0.6%–99.4%, female victimization: 0%–99.0%.

22. See tables at http://www.springerpub.com/content/journals/PA-KnowledgeBase-41410.pdf

23. Nicholls, T. L., Hilterman, E., Tengrstrom, A. (2010). *Decision-making in Abusive Relationships Interview (DIARI)*. Consultation Version 1.0. Port Coquitlam, BC: Forensic Psychiatric Services Commission.

24. However, it has been found that SPJ measures are not necessarily unduly resource intensive (Doyle, Lewis, & Brisbane, 2008; Webster, Martin, Brink, Nicholls, & Desmarais, 2009).

REFERENCES

References marked with an asterisk indicate studies summarized in the online-only tables available at http://dx.doi.org/10.1891/1946-6560.4.1.e15

American Psychological Association. (2002). *American Psychological Association's ethical principles of psychologists and code of conduct*. Retrieved from the American Psychological Association Website: http://www.apa.org/ethics/code/index.aspx

American Psychological Association. (2010). *American Psychological Association's ethical principles of psychologists and code of conduct*. Retrieved from the American Psychological Association Website: http://www.apa.org/ethics/code/index.aspx

Andrews, D. A. (2012). The Risk-Need-Responsivity (RNR) Model of correctional assessment and treatment. In J. A. Dvoskin, J. L. Skeem, R. W. Novaco, & K. S. Douglas (Eds.), *Using social science to reduce violent offending* (pp. 127–156). New York, NY: Oxford University Press.

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

Andrews, D. A., & Bonta, J. (2000). *Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2001). *Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*, 19–52.

Andrews, D. A., Bonta, J., & Wormwith, J. S. (1995). *Level of Service Inventory-Ontario Revision (LSI-OR): Interview and scoring guide*. Toronto, Canada: Ontario Ministry of the Solicitor General and Correctional Services.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2010a). The Level of Service (LS) Assessment of adults and older adolescents. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 199–225). New York, NY: Routledge/Taylor & Francis Group.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2010b). *The Level of Service/Risk-Need-Responsivity (LS/RNR)*. Toronto, Canada: Multi-Health Systems.

Au, A., Cheung, G., Kropp, R., Yuk-Chung, C., Lam, G. L. T., & Sung, P. (2008). A preliminary validation of the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) in Hong Kong. *Journal of Family Violence*, *23*, 727–735.

Babcock, J. C., Green, C. E., & Robie, C. (2004). Does batterer treatment work? A meta-analytic review of domestic violence treatment outcome research. *Clinical Psychology Review*, *23*, 1023–1053.

Babor, T. F., de la Fuente, J. R., Saunders, J., & Grant, M. (1992). *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in primary health care*. Geneva, Switzerland: World Health Organization.

Beck, C. J. A., Walsh, M. E., Mechanic, M. B., Figueredo, A. J., & Chen, M. (2011). *Intimate partner abuse in divorce mediation: Outcomes from a long-term multi-cultural study*. Retrieved from the National Criminal Justice Reference Service Website: https://www.ncjrs.gov/App/publications/Abstract.aspx?id=258888

Beeby, D. (2012, December 23). Major Justice Canada study puts big price tag on spousal violence. *CTV News*. Retrieved from: http://www.ctvnews.ca/canada/major-justice-canada-study-puts-big-price-tag-on-spousal-violence-1.1090000

*Belfrage, H., & Strand, S. (2008). Structured spousal violence risk assessment: Combining risk factors and victim vulnerability factors. *International Journal of Forensic Mental Health*, 7(1), 39–46.

*Belfrage, H., Strand, S., Storey, J. E., Gibas, A. L., Kropp, P. R., & Hart, S. D. (2011). Assessment and management of risk for intimate partner violence by police officers using the spousal assault risk assessment guide. *Law and Human Behavior*. Advance online publication. http://dx.doi.org/10.1007/s10979-011-9278-0

*Bell, M. E., Cattaneo, L. B., Goodman, L. A., & Dutton, M. A. (2008). Assessing the risk of future psychological abuse: Predicting the accuracy of battered women's predictions. *Journal of Family Violence*, 23(2), 69–80.

*Berk, R. A., He, Y., & Sorenson, S. B. (2005). Developing a practical forecasting screener for domestic violence incidents. *Evaluation Review*, 29(4), 358–383.

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, 123, 123–142.

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines and training. *American Psychologist*, 51, 945–956.

Bowen, E. (2011). An overview of partner violence risk assessment and the potential role of female victim appraisals. *Aggression & Violent Behavior*, 16(3), 214–226.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth.

Campbell, J. C. (1986). Nursing assessment for risk of homicide with battered women. *Advances in Nursing Science*, 8(4), 36–51.

Campbell, J. C. (1995). Prediction of homicide of and by battered women. In J. C. Campbell (Ed.), *Assessing dangerousness: Violence by sexual offenders, batterers and child abusers* (pp. 96–113). Thousand Oaks, CA: Sage.

Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M. A., & Laughon, K. (2003). Risk factors for femicide in abusive relationships: Results from a multisite case control study. *American Journal of Public Health*, 93(7), 1089–1097.

*Campbell, J. C., Webster, D. W., & Glass, N. (2009). The Danger Assessment: Validation of a lethality risk assessment instrument for intimate partner femicide. *Journal of Interpersonal Violence*, 24(4), 653–674.

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

*Cattaneo, L. B., Bell, M. E., Goodman, L. A., & Dutton, M. A. (2007). Intimate partner violence victims' accuracy in assessing their risk of reabuse. *Journal of Family Violence*, 22(6), 429–440.

Cattaneo, L. B., & Goodman, L. A. (2005). Risk factors for reabuse in intimate partner violence: A cross-disciplinary critical review. *Trauma, Violence, & Abuse*, 6(2), 141–175.

*Cattaneo, L. B., & Goodman, L. A. (2003). Victim-reported risk factors for continued abusive behavior: Assessing the dangerousness of arrested batterers. *Journal of Community Psychology*, *31*(4), 349–369.

Cavanaugh, M. M., & Gelles, R. J. (2005). The utility of male domestic violence offender typologies. *Journal of Interpersonal Violence*, *20*(2), 155–166.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing inter-rater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127–137.

*Clift, R. J. W., Thomas, L. A., & Dutton, D. G. (2005). Two-year reliability of the propensity for abusiveness scale. *Journal of Family Violence*, *20*, 231–234.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., . . . Rogers, R. D. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, *77*(2), 337–348.

*Connor-Smith, J. K., Henning, K., Moore, S., & Holdford, R. (2011). Risk assessments by female victims of intimate partner violence: Predictors of risk perceptions and comparison to an actuarial measure. *Journal of Interpersonal Violence*, *26*(12), 2517–2550.

Day, T. (1995). When love turns to violence: What you should know about adult abuse. *Journal of American Chiropractic Association*, *32*(9), 31.

Desmarais, S. L., Reeves, K. A., Nicholls, T. L., Telford, R., & Fiebert, M. S. (2012a). Prevalence of physical violence in intimate relationships: Part 1. Rates of male and female victimization. *Partner Abuse*, *3*(2), 140–169.

Desmarais, S. L., Reeves, K. A., Nicholls, T. L., Telford, R., & Fiebert, M. S. (2012b). Prevalence of physical violence in intimate relationships: Part 2. Rates of male and female perpetration. *Partner Abuse*, *3*(2), 170–198.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. (3rd ed.). Thousand Oaks, CA: Sage.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, *29*(5), 617–658.

Doyle, M., Lewis, G., & Brisbane, M. (2008). Implementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatric Bulletin*, *32*, 406–408.

*Dutton, D. G. (1995). A scale for measuring propensity for abusiveness. *Journal of Family Violence*, *10*, 203–221.

Dutton, D. G., & Kropp, P. R. (2000). A review of domestic violence risk instruments. *Trauma, Violence, & Abuse*, *1*(2), 171–181.

*Dutton, D. G., Landolt, M. A., Starzomski, A., & Bodnarchuk, M. (2001). Validation of the Propensity for Abusiveness Scale in diverse male populations. *Journal of Family Violence*, *16*(1), 59–73.

Dutton, M. A. (1996). Battered women's strategic responses to violence: The role of context. In J. Edelson & Z. Eisikovitz (Eds.), *Future interventions with battered women and their families* (pp. 105–124). Thousand Oaks, CA: Sage.

Dutton, M. A., & Dionne, D. (1991). Counseling and shelter for battered women. In M. Steinman (Ed.), *Woman battering: Policy responses* (pp. 113–130). Cincinnati, OH: Anderson.

Dvoskin, J. A., Skeem, J. L., Novaco, R. W., & Douglas, K. S. (2012). What if psychology redesigned the criminal justice system? In J. A. Dvoskin, J. L. Skeem, R. W. Novaco, & K. S. Douglas (Eds.), *Using social science to reduce violent offending* (pp. 291–302). New York, NY: Oxford University Press.

Echeburúa, E., Fernández-Montalvo, J., & Corral, P. (2008). ¿Hay diferencias entre la violencia grave y la violencia menos grave contra la pareja?: Un análisis comparativo [Are there differences between severe violence and non-severe violence against one's partner?: A comparative analysis]. *International Journal of Clinical and Health Psychology*, *8*, 355–382.

*Echeburúa, E., Fernández-Montalvo, J., de Corral, P., & López-Goñi, J. J. (2009). Assessing risk markers in intimate partner femicide and severe violence: A new assessment instrument. *Journal of Interpersonal Violence*, *24*(6), 925–939.

Edens, J. F., & Vincent, G. M. (2008). Juvenile psychopathy: A clinical construct in need of restraint? *Journal of Forensic Psychology Practice*, *8*(2), 186–197.

*Eke, A. W., Hilton, N. Z., Harris, G. T., Rice, M. E., & Houghton, R. E. (2011). Intimate partner homicide: Risk assessment and prospects for prediction. *Journal of Family Violence*, *26*, 211–216.

Ellis, D., & Stuckless, N. (2006a). Domestic violence, DOVE and divorce mediation. *Family Court Review*, *44*(4), 658–671.

*Ellis, D., & Stuckless, N. (2006b). Separation, domestic violence, and divorce mediation. *Conflict Resolution Quarterly*, *23*(4), 461–485.

Feld, S. L., & Straus, M. A. (1989). Escalation and desistence of wife assault in marriage. *Criminology*, *27*(1), 141–162.

Fleiss, J. L., Williams, J. B., & Dubro, A. F. (1986). The logistic regression analysis of psychiatric data. *Journal of Psychiatric Research*, *20*(3), 195–209.

Follingstad, D. R., Laughlin, J. E., Polek, D. S., Rutledge, L. L., & Hause, E. S. (1991). Identification of patterns of wife abuse. *Journal of Interpersonal Violence*, *6*(2), 197–204.

Gelles, R., & Tolman, R. (1998). *The Kingston Screening Instrument for Domestic Violence (KSID)*. Unpublished risk instrument. University of Rhode Island, Providence.

*Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior*, *31*(2), 150–181.

*Glass, N., Laughon, K., Rutto, C., Bevacqua, J., & Campbell, J. C. (2008). Young adult intimate partner femicide—An exploratory study. *Homicide Studies*, *12*(2), 177–187.

*Glass, N., Perrin, N., Hanson, G., Bloom, T., Gardner, E., & Campbell, J. C. (2008). Risk for reassault in abusive female same-sex relationships. *American Journal of Public Health*, *98*(6), 1021–1027.

Gondolf, E. W. (1997). Patterns of reassault in batterer programs. *Violence and Victims*, *12*(4), 373–387.

*Gondolf, E. W., & Wernik, H. (2009). Clinician ratings of batterer treatment behaviors in predicting reassault. *Journal of Interpersonal Violence*, *24*(11), 1792–1815.

*Goodman, L. A., Dutton, M. A., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers—Use of the danger assessment scale in the criminal justice system. *Journal of Interpersonal Violence*, *15*(1), 63–74.

*Grann, M., & Wedin, I. (2002). Risk factors for recidivism among spousal assault and spousal homicide offenders. *Psychology, Crime & Law*, *8*(1), 5–23.

Greaves, L., Hankivisky, O., & Kingston-Riechers, J. (1995). *Selected estimated of the costs of violence against women*. London, Canada: Centre for Research on Violence Against Women and Children.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293–323.

Guo, B., & Harstall, C. (2008). *Spousal violence against women: Preventing recurrence*. Edmonton, AB: Institute of Health Economics.

Guy, L. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey* (Unpublished dissertation). Simon Fraser University, Burnaby, BC, Canada.

Hanson, R. K., Helmus, L,. & Bourgon, G. (2007). The validity of risk assessments for intimate partner violence: A meta-analysis. Retrieved from the Public Safety Canada Website: http://www.publicsafety.gc.ca/res/cor/rep/vra_ipv_200707-eng.aspx

*Hanson, R. K., & Wallace-Capretta, S. (2004). Predictors of criminal recidivism among male batterers. *Psychology Crime & Law*, *10*(4), 413–427.

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised manual*. North Tonawanda, NY: Multi-Health Systems.

Hare, R. D. (2003). *Hare Psychopathy Checklist-Revised* (2nd ed.). Toronto, Canada: Multi-Health Systems.

Hare, R. D (2008). Hare Psychopathy Checklist-Revised (2nd ed.) (PCL-R). In B. L. Cutler (Ed.), *Encyclopedia of psychology and law: Volume I* (pp. 348–350). Thousand Oaks, CA: Sage.

Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior*, *15*(6), 625–637.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, *24*(4), 315–335.

Hart, S. D. (1994). Lethality and dangerousness assessments. *Violence Update*, *4*(10), 7–8.

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, *3*, 121–137.

Hart, S. D. (2009). *Domestic violence*. Invited talk. School of Population and Public Health. University of British Columbia.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare PCL: Screening Version*. North Tonawanda, NY: Multi-Health Systems.

Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. In P. Sturmey & M. McMurran (Eds.), *Forensic case formulation*. Chichester, United Kingdom: Wiley-Blackwell.

Hart, S. D., & Watt, K. A. (2008). Danger Assessment instrument (DA). In B. L. Cutler (Ed.), *Encyclopedia of psychology and law: Volume I* (pp. 185–186). Thousand Oaks, CA: Sage.

*Heckert, D. A., & Gondolf, E. W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence*, *19*(7), 778–800.

Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York, NY: Kluwer.

Heilbrun, K. (2009). *Evaluation for risk of violence in adults*. New York, NY: Oxford University Press.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 1–17). New York, NY: Routledge/Taylor & Francis.

Henderson, A. J. Z., Bartholomew, K., & Dutton, D. G. (1997). He loves me; he loves me not: Attachment and separation resolution of abused women. *Journal of Family Violence*, *12*(2), 169–191.

*Hendricks, B., Werner, T., Shipway, L., & Turinetti, G. J. (2006). Recidivism among spousal abusers: Predictions and program evaluation. *Journal of Interpersonal Violence*, *21*, 703–716.

Herve, H., & Yuille, J. C. (2007). *The psychopath: Theory, research, and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hilterman, E. (2012, April). DIARI assessment in relation with quality of life and revictimization of IPV victims in Barcelona, Spain. In T. Nicholls (Chair), *Understanding and supporting women's decision-making in abusive relationships: The DIARI*. Symposium conducted at the 12th annual conference of the International Association of Forensic Mental Health Services, Miami, FL.

*Hilton, N. Z., & Harris, G. T. (2009). How nonrecidivism affects predictive accuracy: Evidence from a cross-validation of the Ontario Domestic Assault Risk Assessment (ODARA). *Journal of Interpersonal Violence*, *24*, 326–337.

*Hilton, N. Z., Harris, G. T., & Holder, N. (2008). Actuarial assessment of violence risk in hospital-based partner assault clinics. *Canadian Journal of Nursing Research*, *40*(4), 56–70.

*Hilton, N. Z., Harris, G. T., Popham, S., & Lang, C. (2010). Risk assessment among incarcerated male domestic violence offenders. *Criminal Justice and Behavior*, *37*, 815–832.

*Hilton, N. Z., Harris, G. T., & Rice, M. E. (2001). Predicting violence by serious wife assaulters. *Journal of Interpersonal Violence*, *16*(5), 408–423.

*Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal Guide. *Law and Human Behavior*, *32*, 150–163.

*Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment*, *16*, 267–275.

Hisashima, J. (2008). *Validation study of the Domestic Violence Screening Instrument (DVSI)*. Honolulu, Hawaii: Hawaii State Department of Health.

Holtzworth-Munroe, A., & Stuart, G. L. (1994). Treatment of marital violence. In L. Van de Creek, S. Knapp, & T. L. Jackson (Eds.), *Innovations in clinical practice: A sourcebook* (pp. 5–19). Sarasota, FL: Professional Resource Press/Professional Resource Exchange.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York, NY: Wiley.

Kraemer, H., Kazdin, A., Offord, D., Kessler, R., Jensen, P., & Kupfer, D. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, *54*, 337–343.

Kropp, P. R., & Gibas, A. (2010). The Spousal Assault Risk Assessment Guide (SARA). In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 227–250). New York, NY: Routledge/Taylor & Francis.

*Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, *24*(1), 101–118.

Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *Brief spousal assault form for the evaluation of risk (B-SAFER): User manual*. Vancouver, Canada: Proactive Resolutions.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment Guide*. Vancouver, Canada: British Columbia Institute Against Family Violence.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide* (2nd ed.). Vancouver, Canada: British Columbia Institute Against Family Violence.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Spousal Assault Risk Assessment: User's Guide*. Toronto, Canada: Multi-Health Systems.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (2008). *Manual for the Spousal Assault Risk Assessment Guide* (2nd ed.). Vancouver, BC, Canada: The British Columbia Institute Against Family Violence.

Lempert, L. B. (1996). Women's strategies for survival: Developing agency in abusive relationships. *Journal of Family Violence*, *11*(3), 269–289.

Litwack, T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal and clinical considerations. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (2nd ed., pp. 171–271). Hoboken, NJ: John Wiley & Sons.

Maden, A. (2007). *Treating violence: A guide to risk management in mental health*. Oxford, United Kingdom: Oxford University Press.

Magdol, L., Moffitt, T. E., Caspi, A., & Silva, P. A. (1998). Developmental antecedents of partner abuse: A prospective-longitudinal study. *Journal of Abnormal Psychology*, *107*, 375–389.

Marshall, L. L. (1992). Development of the severity of violence against women scales. *Journal of Family Violence*, *7*(2), 103–121.

*McFarlane, J., Campbell, J. C., Sharps, P., & Watson, K. (2002). Abuse during pregnancy and femicide: Urgent implications for women's health. *Obstetrics & Gynecology*, *100*(1), 27–36.

*McFarlane, J., Parker, B., & Soeken, K. (1995). Abuse during pregnancy—Frequency, severity, perpetrator, and risk-factors of homicide. *Public Health Nursing*, *12*(5), 284–289.

*McFarlane, J., Soeken, K., Campbell, J., Parker, B., Reel, S., & Silva, C. (1998). Severity of abuse to pregnant women and associated gun access of the perpetrator. *Public Health Nursing*, *15*(3), 201–206.

Meehl, P. E. (1954/1996). *Clinical versus statistical prediction: A theoretical analysis and review of the literature*. Northvale, NJ: Jason Aronson. (Original work published 1954)

Mills, J. F., Kroner, D. F., & Morgan, R. D. (2011). *Clinician's guide to violence risk assessment*. New York, NY: Guilford Press.

Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Beverly Hills, CA: Sage.

Monahan, J., Steadman, H .J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., . . . Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford University Press.

Mossman, D. (1994). Further comments on portraying the accuracy of violence predictions. *Law & Human Behavior*, *18*(5), 587–593.

Mossman, D., & Somoza, E. (1991). ROC curves, test accuracy, and the description of diagnostic tests. *Journal of Neuropsychiatry and Clinical Neurosciences*, *3*, 330–333.

*Murphy, C. M., Morrel, T. M., Elliott, J. D., & Neavins, T. M. (2003). A prognostic indicator scale for the treatment of partner abuse perpetrators. *Journal of Interpersonal Violence*, *18*(9), 1087–1105.

Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior*, *32*, 352–362.

Nicholls, T. L., Desmarais, S. L., Douglas, K. S., & Kropp, P. R. (2007). Violence risk assessments with perpetrators of intimate partner abuse. In J. Hamel & T. L. Nicholls (Eds.), *Family interventions in domestic violence* (pp. 275–302).New York, NY: Springer Publishing.

Nicholls, T. L., Hilterman, E., & Tengström, A. (2010). *Decision-making in Abusive Relationships Interview (DIARI)*. Consultation Version 1.0. Port Coquitlam, BC: Forensic Psychiatric Services Commission.

Norman, G. R., & Streiner, D. L. (2008). *Biostatistics: The bare essentials* (3rd ed.). Hamilton, Canada: BC Decker.

Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines.* Ottawa, Canada: Ministry of Supplies and Services Canada.

Ontario Ministry of the Solicitor General. (2000, April). *A guide to the domestic violence supplementary report form*. Toronto, Canada: Police Services Division.

Otto, R. K., & Douglas, K. S. (2010). *Handbook of violence risk assessment.* New York, NY: Routledge/Taylor & Francis.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Reeves, C. A., & O'Leary-Kelly, A. M. (2007). The effects and costs of intimate partner violence for work organizations. *Journal of Interpersonal Violence*, *22*(3), 327–344.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63*, 737–748.

Rice, M. E., & Harris, G. T. (1997). Cross validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior*, *21*, 231–241.

Rice, M. E. & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, *29*(5), 615–620.

Rice, M. E., Harris, G. T., & Hilton, N. Z. (2010). The Violence Risk Appraisal Guide and Sex Offender Risk Appraisal Guide for violence risk assessment and the Ontario Domestic Assault Risk Assessment and Domestic Violence Risk Appraisal Guide for wife assault risk assessment. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 99–120). New York, NY: Routledge/ Taylor & Francis.

Rice, M. E., Harris, G. T., Lang, C., & Bell, V. (1990). Recidivism among male insanity acquittees. *Journal of Psychiatry and Law*, *18*, 379–403.

Risk Management Authority. (2006). *Standards and guidelines for risk assessment*. Retrieved from Scotland Risk Management Authority Website: http://www.rmascotland. gov.uk/files/8512/7305/9343/riskAssessmentGuidelines.pdf

Roehl, J., O'Sullivan, C., Webster, D., & Campbell, J. (2005). *Intimate partner violence risk assessment validation study: The RAVE Study practitioner summary and recommendations: Validation of tools for assessing risk from violent intimate partners* (NCJRS

209731). Washington, DC: U.S. Department of Justice. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/209732.pdf

Salekin, T. R., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, *3*(3), 203–215.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499–513.

Singh, J. P., Serper, M., Reinharth, J., & Fazel, S. (2011). Structured assessment of violence risk in schizophrenia and other psychiatric disorders: A systematic review of the validity, reliability, and item content of 10 available instruments. *Schizophrenia Bulletin*, *37*(5), 899–912.

Skeem, J. (2005, March). Identifying patients at risk for repeated involvement in violence: Self perceptions perform better than systematic scales. In J. Skeem (Chair), *Identifying high risk patients and isolating their dynamic risk factors for violence*. Symposium conducted at the American Psychology Law Society conference, San Diego, CA.

*Snider, C., Webster, D., O'Sullivan, C. S., & Campbell, J. (2009). Intimate partner violence: Development of a brief risk assessment for the emergency department. *Academic Emergency Medicine*, *16*, 1208–1216.

Snow Jones, A. S., & Gondolf, E. W. (2001). Time-varying risk factors for reassault among batterer program participants. *Journal of Family Violence*, *16*(4), 345–359.

Spidel, A., Vincent, G., Huss, M. T., Winters, J., Thomas, L., & Dutton, D. (2007). The psychopathic batterer: Subtyping perpetrators of domestic violence. In H. Herve & J. C. Yuille (Eds.), *The psychopath: Theory, research, and practice* (pp. 327–342). Mahwah, NJ: Lawrence Erlbaum Associates.

Stith, S. M., Rosen, K. H., Middleton, K. A., Busch, A. L., Lundeberg, K., & Carlton, R. P. (2000). The intergenerational transmission of spouse abuse: A meta-analysis. *Journal of Marriage and Family*, *62*(3), 640–654.

Storey, J. E., Gibas, A. L., Reeves, K. A., & Hart, S. D. (2011). Evaluation of a violence risk (threat) assessment training program for police and other criminal justice professionals. *Criminal Justice and Behavior*, *38*(6), 554–564.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The conflict tactics (CT) scales. *Journal of Marriage and the Family*, *41*, 75–88.

Straus, M. A. (1990). New scoring methods for violence and new norms for the Conflict Tactics Scales. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families* (pp. 341–367). New Brunswick, NJ: Transaction Books.

Straus, M. A., & Douglas, E. M. (2004). A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence and Victims*, *19*, 507–520.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The Revised Conflict Tactics Scales (CTS-2). *Journal of Family Issues*, *7*, 283–316.

Stuart, E. P., & Campbell, J. C. (1989). Assessment of patterns of dangerousness with battered women. *Issues in Mental Health Nursing*, *10*(3–4), 245–260.

Sullivan, T. P., McPartland, T. S., Armeli, S., Jaquier, V., & Tennen, H. (2012). Is it the exception or the rule? Daily co-occurrence of physical, sexual and psychological partner violence in a 90-day study of substance-using, community women. *Psychology of Violence*, *2*(2), 154–164.

*Thijssen, J., & de Ruiter, C. (2011). Identifying subtypes of spousal assaulters using the B-SAFER. *Journal of Interpersonal Violence*, *26*(7), 1307–1321.

Tolman, R. M. (1989). The development of a measure of psychological maltreatment of women by their male partners. *Violence and Victims*, *4*(3), 159–177.

Walby, S. (2004). The cost of domestic violence. *Women & equality unit*. Retrieved from Devon County Council Website: http://www.devon.gov.uk/cost_of_dv_report_sept04.pdf

Walker, L. E. A. (1984). *The battered woman syndrome*. New York, NY: Springer Publishing.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence* (version 2). Burnaby, Canada: Mental Health, Law & Policy Institute, Simon Fraser University.

Webster, C. D., & Hucker, S. H. (2007). *Violence risk assessment and management*. Chichester, United Kingdom: Wiley.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START)* (Version 1.1). Port Coquitlam, BC: Forensic Psychiatric Services Commission and Odessa, FL: St. Joseph's Healthcare.

*Weisz, A. N., Tolman, R. M., & Saunders, D. G. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence*, *15*(1), 75–90.

*Williams, K. R. (2012). Family violence risk assessment: A predictive cross-validation study of the Domestic Violence Screening Instrument-Revised (DVSI-R). *Law and Human Behavior*, *36*(2), 120–129.

*Williams, K. R., & Grant, S. R. (2006). Empirically examining the risk of intimate partner violence: The Revised Domestic Violence Screening Instrument (DVSI-R). *Public Health Reports*, *121*, 400–408.

*Williams, K. R., & Houghton, A. B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior*, *28*(4), 437–455.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, *136*(5), 740–767.

Special Circumstances. The second author, M. Pritchard, used portions of this project for her Capstone project to fulfill the requirements of her Master's of Public Health degree at Simon Fraser University, Burnaby, BC, Canada.

Correspondence regarding this article should be directed to Tonia L. Nicholls, PhD, Associate Professor, University of British Columbia, Forensic Psychiatric Hospital, 70 Colony Farm Road, Port Coquitlam, BC V3E 2Y2. E-mail: tnicholls@forensic.bc.ca