An overview of partner violence risk assessment and the potential role of female victim risk appraisals ☆

Erica Bowen

Department of Psychology, James Starley Building, Coventry University, Priory Street, Coventry, England, CV1 5FB, United Kingdom

ARTICLE INFO

ABSTRACT

The requirement for accurate appraisals of risk for intimate partner violence has increased with the implementation of pro-arrest policies in the United Kingdom and elsewhere. During the last 10 years, there has been some progress made in terms of the development of actuarial risk assessment instruments, thus providing alternatives to the previously available structured professional judgment approach. In light of these developments, practitioners need information regarding the reliability and validity of such approaches. In addition, research highlights the potential validity of victim appraisals of risk. The aim of this paper, therefore, is to review the existing literature regarding the practice of risk assessment in this field, with emphasis placed on the validity of currently available risk assessment tools, as well as the predictive validity of victim's own appraisal of this risk. Directions for future research are examined along with the implications of the current evidence base for risk assessment practice.

© 2011 Elsevier Ltd. All rights reserved.

Contents

1. Introduction

Intimate partner violence (IPV) represents a constellation of physical, verbal, psychological, emotional, sexual, and financial abuses that occur within the context of a current or former intimate relationship (Home Office, 2005). Although official British statistics suggest that the rate of IPV is falling, it is still estimated to account for 15% of all violent crime (Walby & Allen, 2004). Details from victim surveys around the world testify that IPV is an ongoing, serious, and global issue for women and men.

A recent multi-national survey conducted on behalf of the World Health Organization identified 'current' prevalence rates of women's victimization (experienced in the last 12 months) ranging from between 15% and 71% across 24,097 respondents in ten different countries including Bangladesh, Brazil, Ethiopia, Japan, Namibia, Peru, Samoa, Serbia and Montenegro, Thailand, and the United Republic of Tanzania (Garcia-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006). Of the six European surveys reported between 1992 and 2003, Kury, Obergfell-Fuchs, and Woessner (2004) found that IPV rates varied as a function of how IPV was operationalized and measured, the samples used and the time-frames employed. Consequently, prevalence estimates ranged from 18.7% (identified from self reports using the Conflict Tactics Scale (CTS, Straus, 1979) during last five years: West Germany, 1992), 16.6% (CTS during last five years: East Germany, 1992), 2.4% (lifetime experience of partner violence: Bavaria, 2002); 8–20% depending on age (lifetime experience, female only sample: Sweden, 1999); and 12.9% (CTS experience during last year: Spain, 2003). British data collected in 2007–2008 for the British Crime Survey, indicated that since the age of 16, 27% of women and 17% of men had experienced partner abuse. In the previous 12 months, 5% of women and 4% of men reported such victimization (44% and 32% respectively; Povey, Coleman, Kaiza, & Roe, 2009). Overall, therefore, current international prevalence estimates ranging from 3 to 71% depending on the country, the definition of IPV adopted, and the timeframe employed.

In England and Wales, as in the USA, during the last 10 years there has been increasing emphasis placed on mandatory and victimless arrest and prosecution policies, and specialist court services have emerged in order to more sensitively process cases of IPV (Bowen, 2011). These legislative and policy changes have resulted unsurprisingly in an increase in the number of perpetrators (predominantly men) of IPV who have been sentenced to both custodial and community sanctions. Within this context the ability to accurately determine the likelihood that an individual will re-offend is vitally important; as such appraisals inform all sentence planning activities at all stages of the British criminal justice process. Moreover, with increasing pressure to provide rehabilitation programs for IPV perpetrators informed by the principles of risk, need and responsivity (Andrews & Bonta, 2006), judgments of risk are used to determine which intervention program an individual should be referred to (Bowen, 2011; Hilton, Harris, & Rice, 2010). Consequently, such judgments carry with them considerable responsibility for public safety as well as the human and civil rights of the offender (Hilton & Harris, 2007), and have been described as 'the most important judgments society asks clinicians to make' (Elbogen, 2002, p. 591).

During the last 15 years, increasing attention has been paid to identifying empirically valid risk factors, and developing risk assessment instruments with a view to aiding clinician accuracy when making risk judgments, albeit at a slower rate than has been observed in the general

violence, crime and sexual offending arenas (Kropp, 2004). Dutton and Kropp (2000) published a review of domestic violence risk instruments, which detailed the only two IPV specific risk assessments available in the published literature at that time, the Danger Assessment (Campbell, 1995) and the Spousal Assault Risk Assessment (Kropp, Hart, Webster & Eaves, 1999). In their conclusion, the authors observed that until the late 1990s there were few guidelines for practitioners regarding IPV risk assessment, but that 'this has changed with the proliferation of spousal assault risk assessment instruments in production' (pg. 178). A decade later, it seems timely to review the IPV risk assessment literature in order to determine the amount of progress made. To this end, therefore, the present review has three main aims. First, the main approaches to risk assessment and the formal IPV risk assessment instruments published in the international literature are described. Second, the empirical evidence regarding the reliability and validity of these instruments is reviewed. Finally, the potential role and contribution of victim appraisals to risk assessment is examined with reference to formal assessments, and the empirical literature that explores the validity of victim risk predictions.

2. Literature search methodology

A systematic search of five electronic databases (ASSIA, PsycINFO, Medline, Academic Search Premier, and Scopus) was conducted using all possible combinations of the following keywords and phrases: 'risk assessment', 'risk prediction', 'risk judgment', and 'clinical prediction' were combined with 'intimate partner', 'wife', 'spouse' and 'domestic' which were combined with 'aggression', 'assault', 'violence', and 'abuse'. In addition, 'victim', 'battered women' and 'survivor' were combined with 'risk factors', and 'prediction'. A separate search was conducted for literature pertaining to the use of specific risk assessment instruments, through using the instrument names and acronyms: 'Spousal Assault Risk Assessment', 'SARA', 'Danger Assessment', 'DA', Domestic Violence Risk Assessment Guide', 'DVRAG', 'Ontario Domestic Assault Risk Assessment', 'ODARA', 'Brief Spousal Assault 'B-SAFER' and 'Domestic Violence Screening Instrument', 'DVSI'. The inclusion criteria employed were: published in the English language, published in peer reviewed journals and adult (18+) samples. The publication date range was not limited. Finally, all relevant abstracts were examined, duplicates removed and the reference lists consulted in order to identify additional sources not returned by the electronic database search. Citation reports were also examined to identify relevant articles published subsequently to those already obtained.

3. The nature of risk

It is acknowledged that the concept of risk is shrouded in ambiguity, with little consensus in the empirical and theoretical literature regarding what is meant when we refer to the risk of IPV (Kropp, 2004). Most commonly, studies that examine the risk of IPV recidivism define risk as the *likelihood* of an incident of violence occurring at some point in the future, although studies vary considerably in their operationalization of 'future' and include follow-up periods ranging between months and years. Such a conceptualization of risk places great emphasis on the prediction of behavior without considering the individual context involved.

In practice, the consideration of risk is multi-faceted, and does not simply focus on whether it is likely that an individual will or will not be violent in the future, but also examines the nature of the risk posed (e.g. the type of violence, severity of violence, and imminence), and the circumstances under which such risk may be elevated or reduced (Douglas & Kropp, 2002). However, the definition of risk adopted is likely to vary with the approach to risk assessment undertaken. For example, proponents of the actuarial approach (see Section 3.4 below) will define risk as the statistical probability of an offender being prosecuted for a domestic violence offense during a defined period of time (Hilton et al., 2010). In contrast, individuals who adopt more clinically informed approaches are likely to incorporate a more rounded definition of risk that reflects its multi-dimensional nature. It is also likely that the nature of the risk assessment undertaken, and hence the operationalization of risk that is most germane, will reflect the judicial context and outcomes surrounding the assessment. These issues will be discussed further in Sections 3.2–3.5.

3.1. Approaches to the assessment of risk

Approaches to the risk assessment of IPV reflect the general trends in risk assessment development. That is, assessments broadly fall within three categories: clinical, actuarial, and structured professional judgment approaches (see Douglas, Cox, & Webster, 1999; Moore, 1996; Quinsey, Harris, Rice, & Cormier, 2006 for more comprehensive accounts of this field).

3.2. Unaided clinical IPV risk assessment

Kropp (2004) argues that of these three approaches or 'generations' of risk assessment, unstructured clinical assessment continues to be the most widely used, in part, due to the general lack of structured assessment instruments available to practitioners. Clinical approaches to risk assessment reflect what is referred to as the intuitive form of decision making (Hilton et al., 2010). That is, the use of intuitive or informal procedures to identify relevant information and from which to derive a decision regarding the likelihood of recidivism. This approach requires the clinician to make risk assessment decisions through the idiosyncratic appraisal of information deemed relevant to each individual case and decisions may ultimately reflect personal experience, clinical observations, and intuition (Hilton et al., 2010). Consequently, the nature and scope of information reviewed varies unsystematically across cases (Doyle & Dolan, 2008).

Within the broader violence risk assessment and clinical assessment literatures, this approach has received harsh criticism due to the lack of validity, reliability, and accountability associated with it (Grove & Meehl, 1996). Moreover, the empirical evidence regarding the validity of this approach suggests that clinical risk predictions are only slightly above chance, and that competence in this area varies significantly between clinicians (Monahan, 1981). In going some way to identify the reasons why clinical prediction is so weak, researchers have identified a number of heuristics or biased information processing strategies that are adopted by clinicians in order to make complex decisions. Such biases include illusory correlations — that is, an assumed correlation between two factors that does not exist empirically, for example schizophrenia and high risk of violent crime (Elbogen, 2002). Additional biases may also occur if clinicians fail to examine the base-rates of violence, thereby leading to over or underestimates of risk (Borum, 1996). It has also been suggested that idiosyncratic or highly salient characteristics may unduly inflate perceptions of risk, regardless of whether there is empirical evidence to document an association with violence (Quinsey, 1995). Other biases including confirmatory bias (placing emphasis on evidence which confirms an opinion and paying less attention to evidence to the contrary), hindsight bias (the overesti-

mation of the predictability of an outcome after the fact), and an over-reliance on memory, have also been documented (Borum, Otto, & Golding, 1993). Additional factors associated with accuracy include clinician confidence (McNeil, Sanburg, & Binder, 1998), risk question framing (Monahan, 1996), and clinician and client gender (Coontz, Lidz, & Mulvey, 1994; Lidz, Mulvey, & Gardner, 1993; McNeil & Binder, 1995).

3.3. Actuarial IPV risk assessment

The 'second generation' actuarial approach to risk assessment relies on the use of statistical methods to predict future behavior (Douglas et al., 1999). The emphasis is on predicting a specified behavior within a specified time-frame (Kropp, 2004). In the actuarial assessment of violence risk, predictor items (risk factors) are based on well designed follow-up studies which identify the items that are prospectively associated with the outcome (Hilton et al., 2010). By adopting this method, the optimum selection of items is based on incremental validity — that is, the strongest predictors are chosen, and thereafter only those factors which significantly increase the predictive ability of the model are included. Such approaches may use statistical equations (statistical prediction), actuarial tables and/or algorithmic programs (Grove, Zald, Lebow, Snitz, & Nelson, 2000). The term actuarial is also used to refer to the selection of predictor variables based on the weight of empirical evidence that supports their statistical association with violent outcomes in representative samples (Hilton et al., 2004). Kropp (2004) notes that the aims of actuarial prediction are twofold: to predict violence in a relative sense through comparing the individual to population norms, and to predict violence in an absolute sense by identifying a precise probabilistic estimate of the likelihood of future violence (p. 681).

Ardent debate continues concerning the predictive superiority vs. clinical utility of actuarial assessment methods. Evidence concerning the accuracy of first and second generation risk assessment approaches leans towards second generation assessments as more reliable, although contrary to the expressed opinion of proponents of the second generation approach (e.g., Hilton et al., 2010) the evidence is not completely conclusive. For example, in their meta-analysis of studies that employed both clinical and 'mechanical' (statistical, actuarial, and algorithmic) approaches to either predict human behavior, make medical diagnoses or prognoses, or assess states and traits, Grove et al. (2000) found mixed evidence. Across the 136 studies included in the analysis, it was generally found that mechanical prediction out performed clinical prediction, and was on average approximately 10% more accurate. However, in approximately half of the studies clinical prediction was as accurate as mechanical prediction, and in 6–16% of cases substantially more so. In a meta-analysis confined to comparing the clinical judgment of mental health professionals to actuarial approaches, Ægisdottir et al. (2006) similarly found a 13% greater accuracy for actuarial over clinical judgment approaches. Across the 48 effect sizes examined, just over half (52%) favored statistical prediction, 38% reported no difference between the two approaches and 10% favored the clinical method. Despite this evidence, many clinicians remain reluctant to accept this evidence, and authors caution against the actuarial method being viewed as a panacea (e.g., Grubin, 1997). Indeed, the actuarial approach does have its own limitations.

The first notable limitation is arguably not a limitation if actuarial assessments are used as intended. Douglas et al. (1999) argue that the actuarial approach is limited due to its focus on predicting future behavior, whereas violence risk assessment is argued to be an ongoing process of assessment review and re-assessment through which, in light of a prediction of future violence, a consideration of the possible courses of action to be undertaken to avert future violence arises (Douglas & Kropp, 2002; Hart, 1998). Consequently, while it might be of some use to know the statistical likelihood of re-offending in order

Table 1
Items in the Domestic Violence Screening Instrument, adapted from Williams and Houghton (2004).

DVSI items	DVSI-R items
Prior non-DV convictions	Non-family assaults
Prior assault, harassment, menacing	Family assaults
Prior DV treatment	Prior family violence intervention/treatment
Prior drug or alcohol treatment	Violation of orders/court supervision
History of DV-related restraining orders	Prior current verbal/emotional abuse
History of DV restraining order violation	Frequency of violence in last 6 months
Object used as weapon in commission of crime	Escalation of violence in last 6 months
Children present during domestic violence incident	Substance abuse
Current employment status (unemployment)	Use of objects as weapons
Separation from victim in last 6 months	Children present during prior or current violence
Did victim have restraining order at time of offense	Employment status
Defendant under community supervision at time of offense	

to determine the intensity of supervision required (e.g., whether custodial sentences should be handed down), the approach has little to say about the nature of supervision and risk management strategies that might prove most beneficial. Several other concerns with actuarial approaches have been raised by Hart (1998) and Grubin (1997). These focus on the use of only a limited number of variables from which actuarial tools are derived and which consequently may miss the idiosyncratic nature of individual risk. In addition, it has been argued that the application of aggregate data to individual cases is based on false logic referred to as the 'ecological fallacy' which leads to erroneous assumptions about individuals based on the characteristics of groups to which they belong (Hilton et al., 2010). A final consideration is the extent to which an actuarial assessment is valid when used with a sample or in a setting for which there has been no instrument validation.

In 2004, Kropp observed that the development of an actuarial instrument for IPV which 'yield(s) cut off scores that will allow decision makers, in an absolute sense, to determine risk categories for spousal violence…may never be possible' (p.681). This was based on the belief that IPV risk is not simply a linear function of the number of risk factors present (the fundamental actuarial assumption), but is, to a certain extent, idiosyncratic and context-dependent. It is interesting to note that despite these concerns, three such actuarial risk assessments have since been developed: the Domestic Violence Screening Instrument (DVSI; Williams & Houghton, 2004), the Ontario Domestic Assault Risk

Assessment (ODARA; Hilton et al., 2004), and the Domestic Violence Risk Appraisal Guide (DVRAG; Hilton, Harris, Rice, Houghton, & Eke, 2008).

3.3.1. The Domestic Violence Screening Instrument (DVSI)

The 12-item DVSI was developed by the Colorado Department of Probation Services, and was based on a previously existing 34-item clinical assessment guide. The motivation for scale development came from the local need to increase the speed with which cases were processed. Consequently the information required to complete the DVSI can be drawn from collateral records and databases, and is based on criminal history variables so that the information can be made available to prosecutors soon after a case is opened (Williams & Houghton, 2004). The items are scored either between 0 and 2 or 0 and 3 and the total possible range of scores is 0–30, with higher scores taken to reflect higher 'risk for reoffending, noncompliance with court, and probation orders, and thus, the higher the risk to victims' (pg. 441). A description of the items is in Table 1.

In 2006, a revised version of the DVSI (the DVSI-R) was published (Williams & Grant, 2006). The revision consisted of clarifying a range of items, rewording others and removing items that appeared redundant on the basis of the earlier version. The resulting scale contains 11 items (see Table 1), seven of which relate to prior behavior. In addition, the authors noted that practitioners were reluctant to completely endorse the use of an actuarial tool as they felt that they wanted the ability to express a clinical opinion. Consequently, practitioners are now also able to report the potential imminence of risk to intimate partner and others in two separate items.

3.3.2. The Ontario Domestic Assault Risk Assessment (ODARA)

Developed by the same group of researchers, it is argued that the ODARA and DVRAG form part of an assessment system designed to enable police officers or victim counselors to conduct an actuarial risk assessment (ODARA) which does not require specialist information, and to provide qualified professionals with a more in-depth and accurate actuarial tool (DVRAG; Hilton et al., 2010). These two tools are described below and Table 2 provides a comparison of items and scoring between both instruments.

The ODARA was developed on the rationale that it would be used by police to inform decisions regarding whether to detain a suspect accused of IPV, and whether to offer additional support to victims. Moreover, it was expected that the ODARA would help to inform decisions concerning bail applications, conditions applied to conditional release and sentencing (Hilton et al., 2010). It consists of 13 items that reflect IPV specific risk factors, and also more general risk factors for criminal behavior (Hilton et al., 2004). In each instance, the item is rated

Table 2
ODARA and DVRAG items and scoring, adapted from Hilton et al. (2010).

ODARA		DVRAG		
Item	Scores	Item	Range	Scores
Prior domestic incident	0 or 1	Number of prior domestic incidents	0, 1, ≥2	−1, 0, +5
Prior nondomestic incident	0 or 1	Number of prior nondomestic incidents	0, ≥1	−1, +5
Prior custodial sentence of 30 days or more	0 or 1	Prior custodial sentence of 30 days or more	No, Yes	−1, +2
Failure on prior conditional release	0 or 1	Failure on prior conditional release	No, Yes	−1, +2
Threat to harm or kill at the index assault	0 or 1	Threat to harm or kill at the index assault	No, Yes	0, +1
Confinement of the partner at the index assault	0 or 1	Confinement of the partner at the index assault	No, Yes	0, +1
Victim concern	0 or 1	Victim concern	No, Yes	0, +2
More than one child	0 or 1	Number of children	0–1, ≥2	−1, +1
Victim's biological child from a previous partner	0 or 1	Victim's biological children from a previous partner	0, 1, ≥2	−1, 0, +2
Violence against others	0 or 1	Violence against others	No, Yes	0, +8
Substance abuse	0 or 1	Substance abuse score	≤1, ≥2 factors present	−2, +2
Assault on victim when pregnant	0 or 1	Assault on victim when pregnant	No, Yes	0, +5
Barriers to victim support	0 or 1	Number of barriers to victim support	0, 1, 2 factors	−1, 0, +4
–		Psychopathy Checklist-Revised score	≤9, 10–16, ≥17	−1, +1, +6

as present (1) or absent (0), and consequently an individual's scale score may range from 0 to 13. The authors derived seven score categories based on the distribution of scores and computed the associated recidivism rate during a follow-up period of nearly five years for a sample of 589 Canadian offenders. On this basis 5% of offenders with a score of 1 recidivated during this period, in contrast to 41% of those with a score of 4, and 70% with a score between 7 and 13. In total, 80% of those who recidivated had scores between 1 and 4 on the scale.

3.3.3. Domestic Violence Risk Assessment Guide (DVRAG)

The DVRAG was developed from the ODARA through the addition of clinically relevant information that is not routinely available to the police (Hilton et al., 2008). Such information included scores on other IPV specific, and general violence risk assessments (SARA; DA; DVSI, Williams & Houghton, 2004; PCL-R, Hare, 1991; VRAG, Harris, Rice, & Quinsey, 1993). The sample used to determine the incremental predictive validity of the measure was obtained from police records, and included only men who had committed either a physical assault, or had made a credible death threat with a weapon in hand (regardless of whether they had subsequently been arrested or criminally charged). Using a file review based PCL-R assessment, it was found that PCL-R scores provided the greatest improvement in predictive validity from the basic ODARA scores (ODARA ROC area = .65; DVRAG ROC area = .70, $p < .05$). Consequently, the 14 DVRAG items include all 13 ODARA items along with an item accounting for PCL-R scores. However, rather than the dichotomous scoring adopted for the ODARA, the DVRAG scoring system uses the Nuffield weighting system, consistent with the VRAG (see Table 2).

3.4. Structured professional judgment IPV risk assessment

The 'third generation' structured professional judgment or 'empirically validated structured decision-making' (Douglas et al., 1999) approach facilitates the systematic assessment of a number of specified risk factors through the provision of guidelines, but requires the individual professional (who may or may not be clinically trained, Kropp, 2004) to base their overall judgments of the nature of the risk posed by an individual offender on their clinical understanding of the case (Douglas & Kropp, 2002). The set of identified risk factors is deemed to be the minimum on which an appraisal of risk should be conducted. Consequently this approach has flexibility that is not apparent in purely actuarial approaches, but more rigidity than unstructured clinical judgment. This approach changes the emphasis from one of risk prediction to one of risk management (Doyle & Dolan, 2008). Two assessment tools, the Spousal Assault Risk Assessment (SARA) and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) have been developed by the same team of researchers, drawing on the same underlying model of risk. The items for both assessments are presented in Table 3.

3.4.1. The Spousal Assault Risk Assessment (SARA)

The SARA provides those charged with the responsibility of assessing IPV risk, with a checklist of 20 empirically validated IPV risk factors, along with guidelines for best practice in evidence gathering and evaluation (Kropp, Hart, Webster, & Eaves, 1999). The rationale behind the SARA is that non-clinicians should be able to make use of it in order to validate their decisions, and ensure that defensible risk management decisions are made. To this end, each item is associated directly with one or more risk management strategies, depending on its nature (Dutton & Kropp, 2000).

The first three items of the SARA reflect criminal history. Items 4–10 reflect psychosocial risk factors. Items 11–17 examine specific details of the history of IPV for the individual. The final three items consider details of the index offense. The SARA has been conceptualized as two parts, with part 1 consisting of items 1–10 which are general violence recidivism risk factors, whereas those in part 2 which

Table 3
SARA and B-SAFER items, adapted from Kropp et al. (1999) and Kropp and Hart (2004).

SARA	B-SAFER
Past assault of family members	Serious physical/sexual violence
Past assault of strangers/acquaintances	Serious violent threats, ideation or intent
Past violations of conditional release or community supervision	Escalation of physical/sexual violence or threats/ideation/intent
Recent relationship problems	Violations of civil or criminal court orders
Recent employment problems	Negative attitudes about spousal assault
Victim of and/or witness to family violence during childhood/adolescence	Other serious criminality
Recent substance abuse/dependence	Relationship problems
Recent suicidal or homicidal ideation/intent	Employment and/or financial problems
Recent psychotic and/or manic symptoms	Substance abuse
Personality disorder with anger, impulsivity, or behavioral instability	Mental disorder
Past physical assault	Other considerations
Past sexual assault/sexual jealousy	
Past use of weapons/credible threats of death	
Recent escalation in frequency or severity of assault	
Past violation of no-contact orders	
Extreme minimization or denial of spouse assault history	
Attitudes that support or condone spousal assault	
Severe and/or sexual assault (current/most recent offense)	
Use of weapons and/or credible threats of death (current/most recent offense)	
Violation of no contact order (current or most-recent offense)	

Note: Items for both instruments are scored as 0 = absent; 1 = possibly present; and 2 = present.

consists of items 11–20 are deemed to be IPV specific risk factors (Grann & Wedin, 2002).

Each item has a specific evidentiary definition in order to justify the coding of it as either present, sub-threshold or absent. Items are scored on a scale of 2, 1 and 0 depending on the extent to which each item appears to be present from the available evidence (0 = absent and 2 = present). That some of the items reflect clinical constructs may potentially cause problems for non-clinician risk assessors, the SARA allows for the consideration and incorporation of clinical opinion as well as the results of more structured clinical assessments. Kropp et al. (1999) also note that in some instances (e.g., denial/pro-IPV attitudes), additional evidence may be gathered through the use of validated scales (e.g., Psychological Maltreatment of Women Scale; Tolman, 1989), and that perpetrator report should be cross-validated with victim report where available, either by consulting victim statements, or interviewing the victim explicitly. It is also noted that relevant additional items can be included if they are deemed relevant to an individual case. Consequently, the SARA should be considered a starting point for best practice, and not a prescriptive assessment framework.

Although numerical item scores are derived, and total scale scores are possible to calculate, the SARA cannot and must not be used as an actuarial measure except for research purposes (Kropp et al., 1999). In contrast, clinicians should derive broad risk categories as either low, medium or high. Although Kropp et al. note that the SARA assumes a general linear association between the number of items present and risk categorization (the basis of actuarial scales) but the authors also highlight the potential importance of critical items (clinical judgment). These are risk factors that, if present, may be sufficient by themselves to categorize the perpetrator as high risk of imminent harm in the absence of other factors (e.g., access to a weapon/fire arm).

3.4.2. The Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp & Hart, 2004)

The B-SAFER was developed after the SARA, on the basis that the SARA was both time and resource intensive to complete, and that it was less than optimal for use by police personnel who rarely possessed the clinical skills to conduct many of the formal assessments required to complete the SARA items proficiently (e.g., personality disorder, psychopathy, mental illness, substance use; Kropp, Hart, & Belfrage, 2005). Resulting from the factor analysis of SARA items which identified seven underlying dimensions, the B-SAFER comprises ten items, reflecting two sections: spousal assault (items 1–5) and psychosocial adjustment (items 6–10, Table 3). These sections generally require information that can be readily obtained by the offender and victim at the time of the index offense, or through other sources of information available to police officers. Although the scale consists of 10 items, each is appraised with respect to past and current (within the last four weeks) behavior, and consequently 20 ratings are recorded.

From Table 3 it is possible to see the broad overlap between SARA and B-SAFER items. For instance, the 'Negative attitudes about spousal assault' B-SAFER item reflects the combined intentions of the 'Extreme minimization/denial' and 'Attitudes supporting/condoning spouse abuse' SARA items. Although intended to reduce the burden of resources and specialist qualifications required by raters, it is interesting that the coding instructions for the 'mental disorder' item makes reference to the use of the ICD-10 and DSM-IV criteria, thus still emphasizing the requirement for specialist knowledge, although this item can be coded provisionally without access to such information (Kropp & Hart, 2004). More recently it has been suggested that evidence from psychiatric reports can also be used (Kropp et al., 2005). Items are scored as $0 =$ absent, $1 =$ partially present, and $2 =$ present. Total scale scores can be derived from summing the scores for the 'present', 'past' and 'past + present' scales (Au et al., 2008). However, as the main focus is on preventing IPV, the secondary task of the assessor is to determine appropriate risk management strategies based on the identified risk factors (Kropp et al., 2005).

3.5. Reliability and validity of second and third generation risk assessment instruments

The reliability of a measure refers to the consistency with which an instrument measures across multiple administrations and time. As Quinsey et al. (2006) note, within the field of corrections prediction research, reliability is typically assessed when the same tool is used by multiple assessors of the same individual. Validity, in contrast, refers more specifically to the accuracy of the measure, or the extent to which what is being assessed is what is intended. Predictive validity is of most importance when examining the accuracy of tools designed to predict recidivism (Quinsey et al., 2006), and refers to the extent to which individuals who reoffend are identified as likely recidivists prior to their reoffending. Criterion validity is a proxy measure of this conducted to determine the ability of a measure to discriminate between two known groups rather than waiting for a prospective outcome. Predictive validity is typically computed using Relative Operating Characteristics Area Under the Curve (ROC AUC) analysis in which the specificity of a tool (the extent to which individuals who reoffend are predicted to be recidivists, or true positives) is plotted against the sensitivity of the tool (the extent to which those who do not reoffend are inaccurately identified as recidivists, or false alarms/false positives). An AUC of .50 is taken to indicate a predictive accuracy of chance and AUCs are calculated to determine whether they are a significant improvement above this with a maximum AUC of 1.0 which would indicate perfect predictive accuracy (Mossman, 1994).

3.5.1. Domestic Violence Screening Instrument (DVSI)

The DVSI was validated on a sample of 1465 male offenders arrested for domestic violence offenses committed against female partners

(Williams & Houghton, 2004). An 18 month follow-up criminal record review was conducted. In addition, a sample of 125 female partners of these men was obtained. The internal consistency of the DVSI was found to be adequate ($\alpha = .71$). Concurrent validity was examined in relation to Total SARA scores and clinician perceptions of imminent risk to the partner. The resulting correlations ($r = .54$ and .57, respectively) were found to be relatively strong, accounting for just under one third of the variance (29% and 32%, respectively). Discriminant validity was calculated through the correlation between DVSI scores and ratings of the perceived imminent risk to others. As expected, this correlation was substantially lower than that achieved from the concurrent validation ($r = .15$). Based on the criminal conviction review, 18 months post-arrest, it was found that the ROC AUC was .61 for domestic violence reoffending, and .65 for general reoffending, both estimates of which were significantly different from .50 (no relationship).

An interesting pattern of findings emerged from the 6-month post sentence follow up of the 125 victims. It was found that the ability of the DVSI to accurately predict abuse varied as a function of the severity of the behaviors reported. For example, when predicting controlling behaviors, less severe threatening and violent behaviors, the DVSI performed no better than chance (AUC = .58, .56 and .49, respectively). In contrast, when predicting severe threatening behavior and very severe violent behavior the DVSI performed significantly better than chance with adequate predictive validity identified (AUC = .68 and .65, respectively).

In their construction and validation of the DVRAG (see Section 3.5.3) Hilton et al. (2008) found that across the 5 year recidivism period, DVSI scores were significantly and positively associated with recidivism ($r = .17$), number of recidivistic incidents ($r = .24$), the total victim injury ($r = .18$), and the number of severe CTS incidents ($r = .19$). In addition, DVSI scores were also significantly positively associated with DVRAG scores ($r = .49$).

Data from the revised DVSI indicated that the instruments predictive validity increased with the incorporation of additional information from raters. For example, by itself the AUC for the DVSI-R indicated good predictive validity (.71), when predicting repeated arrest during an 8-month follow-up period (Williams & Grant, 2006). However, when propensity scores were calculated from perpetrator, offense and DVSI-R information, it was found that the most accurate model reflected the inclusion of perpetrator characteristics, the victim being an intimate partner, DVSI-R total scores, and clinician appraisals of the imminence of risk to partners and others, which led to a combined AUC of .84. Evidence for the consistency of prediction accuracy was also found through estimating the association between DVSI-R scores and ten randomly selected samples of 1500 cases. It was found that the mean AUC indicated a good level of predictive validity (.70, range .68–.73).

Together these data illustrate that both that the original version of the DVSI exhibits a moderate degree of predictive accuracy, but that this is marginally greater for general offending than for IPV when relying on official reports of recidivism. When victim data are used, however, the DVSI is better at predicting severe threatening and violent behavior than it is less severe and controlling behaviors. It also appeared to be modestly associated with other IPV risk assessment instruments showing construct validity. The revised version seems to be more accurate at predicting IPV, but predictive accuracy is greatest when the DVSI is included within a broader assessment of perpetrator, victim and clinician variables.

3.5.2. ODARA

The ODARA was constructed from data relating to 589 IPV offenders followed up for an average of 4.79 years after the police recorded an alleged domestic incident for which there was evidence of 'forceful physical contact by a man against his current or former wife or common-law wife' (Hilton et al., 2004; p. 269). The index assault need not have led to arrest or charge at that time. In addition a cross validation sample of 100 perpetrators was also identified. It was found on the

construction sample that the predictive accuracy of the ODARA was good, yielding an ROC = .77, equivalent to a large effect size (Cohen's d = 1.1). Concurrent validity was assessed through correlating ODARA scores with those on the SARA ($r = .60$) and the DA ($r = .43$). When cross validation of 100 cases was conducted, it was found that the ROC reduced to .72; moreover, neither the SARA nor the DA significantly predicted recidivism. ODARA scores were also found to be significantly positively associated with victim injury scores ($r = .37$), number of severe IPV incidents ($r = .34$), total number of recidivistic offenses ($r = .37$), and time at risk ($r = -.34$). Although limited, these data provide evidence that the ODARA has very good predictive accuracy, and levels of construct validity akin to those achieved by other risk assessment instruments.

3.5.3. DVRAG

To date the only published validation study for the DVRAG is that detailing its construction (Hilton et al., 2008). The DVRAG was validated on two samples of IPV perpetrators identified by examining police records as individuals who had 'committed an act of physical assault or credible threat of death with weapon in hand in the presence of a victim who was a current or former wife or common-law wife (Hilton et al., 2008). Using a mean follow-up period of 5 years, in the first sample of 303 offenders (the construction sample), it was found that DVRAG scores correlated significantly and positively with recidivism ($r = .30$), number of recidivistic offense ($r = .37$), severe physical abuse (CTS, $r = .37$), and total injury (CTS, $r = .39$). The ROC area in sample 1 suggested good predictive validity (AUC = .71), which was significantly different from .50. In the second sample of 346 offenders, the data from whom had not been used to construct or validate the instrument, similarly good predictive validity was also found (ROC = .70). In addition, the inter-rater reliability for DVRAG items was $r = .92$. These data indicate that the DVRAG has the potential for high levels of predictive accuracy, and similar levels of concurrent validity to other risk assessments. More research is required however to determine its predictive validity across a broader range of participants and time frames.

3.5.4. SARA

Although not an actuarial measure (Kropp et al., 1999), there has been some research conducted to determine the predictive validity of the SARA and inter-rater reliability of the SARA items. Kropp and Hart (2000) detail the reliability and validity of SARA judgments in a sample of 2681 male Canadian offenders obtained from probation (n = 1671) and prison (n = 1010). In general across the items inter-rater reliability was high (>.80). However, the inter-rater reliability for Part 1 total scores and Part 1 items present, and agreement on summary ratings (e.g., Low vs. Moderate vs. High) was less than ideal (0.50–0.68) with the identification of critical items particularly poor (0.18 part 1 and 0.38 part 2). Evidence was found for criterion-groups validity. When ratings assigned to inmates with and without a documented history of IPV were compared, those with an IPV history had significantly higher scores. When comparing the assessments of recidivist and non-recidivist offenders, a series of significant differences were identified on SARA scores. For example, although the total and part 1 scores did not differ significantly between groups ($p < .07$ in both cases) significant differences were found between groups for the total number of factors and the number of factors on part 2, and the total number of critical items and those within part 2, providing more (albeit inconsistent) evidence for the criterion validity of the measure.

Concurrent validity was assessed by examining associations with the PCL-SV, the VRAG (Quinsey et al., 2006), and an actuarial scale for the assessment of general criminality (i.e., the GSIR; Nuffield, 1982). In general, some evidence for the concurrent validity was found when examined in relation to scores on the PCL-SV, but there were no significant associations between SARA total scores and scores on the other two measures. As might be expected, however, part 1 scores which reflect a general violence predisposition were significantly

associated with scores on all measures, as were the number of factors present on part 1.

In a Swedish study, Grann and Wedin (2002) examined the predictive validity of actuarial scale scores and item scores derived from the retrospective file-only SARA assessments of a sample of 88 court referred IPV perpetrators followed up for a period of seven years. It was found that the part 1 and 2 scores performed only marginally better than chance in predicting recidivism during this period. In addition, three items were found to significantly predict recidivism: past violation of conditional release/community supervisions, personality disorder with impulsivity/anger/behavioral instability and extreme denial or minimization of spousal assault history.

Predictive validity analyses using AUC were conducted for 6 months, 1 year, 2 years and 5 year follow-ups. The lowest AUCs were obtained at 6 month assessment .49, .54 and .52 (part 1, part 2 and total, respectively; n = 88), although these data suggest only marginal predictive validity. The highest AUCs were obtained at the 5 year follow-up period .59, .62 and .65 (part 1, part 2 and total, respectively; n = 56), although only the latter two suggest adequate predictive validity of the scale. Of interest, at 1 year follow-up, both the PCL-R and HCR-20 (Historical items only) outperformed the SARA, with predictive validity in the high adequate to good range (AUC .70 and .68 respectively). It is quite possible that the measurement error due to the file-only coding of the SARA items may have impacted on its overall predictive accuracy. Replication of this study with an appropriate SARA, PCL-R, and HCR-20 dataset would be useful to determine the extent to which the SARA is able to uniquely contribute to IPV risk prediction.

Additional data concerning the reliability and validity of the SARA has arisen from the validation studies developed for other risk assessments. For example, Data from the DVSI validation study (Williams & Houghton, 2004) indicated that the SARA part 1 and part 2 items had adequate internal consistency ($\alpha = .66$ and .73 respectively). According to Hilton et al. (2004), the predictive accuracy of the SARA on a sample of 589 IPV perpetrators was ROC .64. Heckert and Gondolf (2004), in their comparison of victim and SARA risk prediction accuracy, found that for their study of 499 female partners using a 15-month follow-up, the predictive accuracy of a simulated SARA assessment (using variables selected from an existing battery that were deemed to reflect SARA items) was only just acceptable (AUC .64). In their construction of the DVRAG, Hilton et al. (2008) considered the potential contribution of adding SARA scores to the ODARA. It was found that in their construction sample of 303 offenders, during the follow-up period (mean 5 years), SARA scores correlated significantly with wife assault ($r = .18$), number of recidivistic incidents ($r = .22$), total victim injury ($r = -.21$) and the number of severe CTS incidents ($r = .20$). It was also significantly positively correlated with ODARA scores ($r = .57$). Of particular importance is the finding that the SARA is better at predicting minor than serious IPV assaults, as indicated by the negative correlation between SARA scores and total victim injury.

Taken together, these studies indicate that the SARA has some validity and reliability although this is perhaps not as clear as might be expected. Moreover, it is of theoretical interest that the SARA has been outperformed by other risk assessments used for the assessment of generic violence risk. However, results of these studies need to be interpreted with caution as in the majority of cases (except Kropp & Hart, 2000), the SARA was not used as intended due to the reliance on file information rather than the comprehensive perpetrator and victim assessment that is recommended. Consequently, it is possible that the actual performance of complete SARA assessments is better, and that the predictive accuracy is greater. Considerably more prospective longitudinal research is required to clarify this issue.

3.5.5. B-SAFER

Au et al. (2008) published a small scale validation study of the B-SAFER using 43 IPV perpetrators and 46 controls identified from their

participation in clinical and family services in Hong Kong. In contrast to controls, IPV men scored significantly higher on ratings of all current items except violation of court orders and mental health problems. Concurrent validity was estimated based on the correlation between B-SAFER subscale scores and scores on the CTS-2 subscales. It was found that the B-SAFER scores correlated positively with both the psychological ($r=.34$, $p<.01$) and physical abuse ($r=.38$, $p<.01$) subscales of the CTS. Again, using the CTS-2, the authors examined the construct validity of the B-SAFER by conducting a discriminant function analysis to determine membership of the two groups (IPV perpetrators vs. Control). It was found that 95% of cases were correctly classified. In addition, results showed B-SAFER current total rather than past total scores were the most important predictor of IPV status, with a fourfold increase in the likelihood of being an IPV perpetrator with each one point increase in current total scores. It is evident from this study, the only published validation study of the B-SAFER, that tentative evidence exists for the validity of the instrument. However, considerably greater attention needs to be paid to determining the predictive validity of this tool among a larger sample drawn from an ethnically diverse community.

3.6. Summary

Overall, each of the instruments exhibits at least a moderate degree of predictive accuracy. Taken together, the data reviewed indicate a general trend towards actuarial instruments out-performing those cast from the structured professional judgment mold. In a review of the performance of these instruments (except the B-SAFER), including the Violence Risk Assessment Guide, Hilton and Harris (2007) calculated mean ROC areas for each from the available literature and found that the DVRAG outperformed all others (ROC = .70, good predictive validity), based on one study, followed by the VRAG (ROC = .68, adequate predictive validity), ODARA (ROC = .67, adequate predictive validity), SARA and DA (ROC = .62, adequate predictive validity) and the DVSI (ROC = .56, marginal predictive validity). These data indicate that only the DVRAG and ODARA achieved at least a medium effect, comparable to Cohen's $d \geq .5$, or ROC = .64 (Rice & Harris, 2005), whereas the DVSI, SARA and DA achieved small effects. In practice the actual observed difference in the performance of the instruments is not large. What is notable from the literature reviewed, however, is the small number of prospective large scale validation studies available, and the restricted number of countries from which they have emerged. Consequently it would be premature to recommend one particular risk assessment tool over and above all others until suitable data exists from which such a decision can be made. What should guide decisions regarding which instrument to use is the context of the risk assessment and the requirements placed on the assessor.

4. Victims as risk assessors

The integration of victim data to inform assessments of risk and dangerousness is not a new concept. In fact it is recommended that such information routinely be drawn on when assessing IPV men in order to obtain a more valid estimate of risk (Kropp, 2007). As can be seen from the items of the ODARA and DVRAG and best practice assessment guidelines, and the instructions for the SARA, the use of victim perceptions and experiences are typically integrated into risk judgments to a more or less explicit extent. Evidence from a study using the SARA lends weight to this recommendation. It was found that when victim-reported information was incorporated into assessments based solely on offender and file information, increases in perceived risk were noted (Whittemore & Kropp, 2002). This suggests that victims provide important and clinically relevant information, however only one risk assessment has been devised which relies solely on such information.

4.1. The Danger Assessment Scale (DAS; Campbell, 1995)

The DAS is an actuarial assessment insofar as it draws upon a retrospective empirical evidence base concerning the risk factors for either intimate partner homicide or severe IPV. It was designed primarily to enable victims of IPV to assess their danger of being seriously injured and/or murdered by a current or former partner, and also to assess the women's risk of killing their abusive male partner (Campbell, Webster, & Glass, 2009). The 20-item measure is split into two sections. The first section requires women to complete a retrospective calendar detailing their experiences of IPV during the past year. The second section is a more formulaic 20-item self report measure determining the presence of a range of risk factors. Each item is scored dichotomously (yes/no), and scoring is based on summing the total number of 'yes' responses, with risk deemed to increase with the increase in positive responses.

4.2. Reliability and validity of the DAS

Goodman, Dutton, and Bennett (2000) examined the predictive validity of the DAS across a three-month follow-up period with a sample of 49 women who attended a domestic violence intake center after the arrest of their partner for suspected IPV. Using logistic regression, it was found that a one standard deviation increase in DAS scores was associated with a four-fold increase in the likelihood of repeat abuse during the follow up period. In their construction of the ODARA, Hilton et al. (2004) examined the predictive accuracy of the DA by means of comparison. It was found that the ROC area for the DA on a sample of 589 IPV perpetrators indicated only marginal predictive validity (.59). However, this assessment was conducted on data that were not obtained directly from the victims themselves and is consequently of questionable validity. Moreover, recidivism was assessed as any subsequent violence against an (ex-)wife or (ex-) partner, regardless of subsequent police action. Consequently, this variable may have included a range of minor incidents as well as those deemed to be serious and/or life threatening.

The utility of the DA in predicting comparatively minor, non-lethal IPV as examined in this study is questionable as this is not the purpose for which the DA was designed. This, therefore, provides perhaps erroneous evidence regarding the predictive validity of the DA. Heckert and Gondolf (2004) found that the predictive accuracy of the DA in their 15 month follow up study was good (AUC = .70) for repeat reassault, based on a simulated DAS in which variables were selected from an item pool based on their perceived reflection of DAS items. Based on a weighted scoring algorithm devised from a multivariate examination of femicide cases (contrasted to abused controls), Campbell et al. (2009) identified four levels of danger which encapsulate DA score ranges. Scores of 0–7 were classified as variable danger; 9–13 increased danger; 14–17, severe danger and scores of 18 and above extreme danger. In order to determine the predictive validity of the 20-item DA, retrospective data were examined from attempted homicide cases (n = 194) as well as abused community based controls (n = 414). When the DA danger levels were compared between attempted femicides and the community control group the AUC for the ROC curve indicated excellent predictive validity (.916). It has also been found that five items from the DAS can be used as a screening instrument for future violence in an emergency department population. Snider, Webster, O'Sullivan, and Campbell (2009) found that by using items relating to the escalation of violence, use of weapon, perception of the perpetrator as capable of killing, violence during pregnancy and violent jealousy, across which three 'yes' responses were provided, a good level of predictive accuracy was identified (AUC = .79) in a sample of 400 women followed up after 9 months.

The main conceptual difficulty surrounding the validation of each version of the DA, has been the reliance on a retrospective approach, whereby known victims of severe IPV or proxies (individuals who

knew a victim well enough to comment, in the case of homicide) identify the presence of a risk factor after the outcome (severe IPV or homicide). It is quite possible, as Campbell et al. (2009) acknowledge that such an approach may increase the likelihood of an individual identifying the presence of a risk factor on the basis of an illusory correlation, (i.e., had the outcome not been known, the risk factor may not have been reported as present). Such a phenomena would result in an inflated association between predictor and criterion variables.

4.3. The ability of victim appraisals of risk to predict future harm

The issue of the extent to which victims can accurately perceive their risk has drawn some attention in the literature during the last decade. Specifically, a small number of studies have examined the ability of female victims of IPV to accurately predict the future likelihood of their own victimization, relative to clinicians and the structured risk assessments already examined. It has been suggested by some, that as these women live their daily lives through minute-by-minute risk assessment, they should have a unique insight into their partners behavior which makes them relative 'experts' in risk assessment within their own interpersonal context (Dutton, 1996). Conversely, it has also been argued that due to the continued exposure to abuse and resulting psychological trauma, female victims are less likely to be able to accurately perceive their own risk (Campbell, 1995), and that their recall of particularly severe episodes may be specifically impaired (Browne, 1987). Moreover, the effects of persistent psychological abuse, name calling and denigration may also render a female victim less confident in the validity of her perceptions and judgment (Dutton & Dionne, 1991).

4.3.1. Accuracy of victim appraisals

Since 2000, six articles were published that have explicitly examined the accuracy of victim predictions (Bell, Bennett Cattaneo, Goodman, & Dutton, 2008; Bennett Cattaneo, Bell, Goodman, & Dutton, 2007; Bennett Cattaneo & Goodman, 2003; Heckert & Gondolf, 2004; Snider et al., 2009; Weisz, Tolman, & Saunders, 2000). Details of these studies are presented in Table 4 below.

As can be derived from Table 4, the overall evidence indicates that victim appraisals of their own risk appear to be consistently valid predictors of future victimization, with approximately two thirds of victims correctly identifying their level of risk (Bell et al., 2008; Bennett Cattaneo et al., 2007); and, in some instances, victim appraisals out-perform the predictive validity of standardized risk assessment tools (Heckert & Gondolf, 2004). In addition, Heckert and Gondolf found that when risk factors (identified from case materials or through using the DA) were combined with characteristics and women's risk appraisals the predictive validity remained within the good range and increased (AUC = .73 and .83 respectively).

The predictive accuracy of victim appraisals appears relatively robust given the methodological variations apparent across studies. For example, each of the studies adopted a different definition of the IPV criterion variable. Weisz et al. (2000) defined 'severe domestic violence' on the basis of severe items in the CTS, which also included threats to kill or threats with a gun or a knife. In contrast Bennett Cattaneo and Goodman (2003) asked survivors whether their partner had contacted them or attempted to contact them when they did not want him to; whether there had been conflict which 'involved physical fighting including grabbing, pushing or throwing things' and whether there had been any sexual contact due to 'force, threats or being afraid of what would happen if you didn't give in' and also whether their partner had 'threatened you with physical harm or destroyed your property' (p. 356). If a survivor provided a positive response to any of these questions it was taken as an indication of continued abuse. Ultimately, although 27.7% of cases were identified as 'continued abuse' the majority of these (26.5%) reflected unwanted contact rather than reassault. Heckert and Gondolf (2004) derived a

reassault variable from cumulative data collected across a 15-month follow-up period. This was derived through telephone administration of a number of questionnaires concerning verbal abuse, threats, physical aggression and controlling behaviors. The outcome of interest was repeat re-assault — that is, individuals whose partners reported experiencing more than one incident that included one of the tactics on the physical aggression subscale of the CTS (p.784). Similar to Bennett Cattaneo and Goodman, Bennett Cattaneo et al. (2007) adopt a broad definition of IPV characterized by the same features, but with these features assessed using a range of standardized assessment instruments. Finally Bell et al. (2008) focused purely on the psychological abuse variable used in the earlier study (Bennett Cattaneo et al., 2007).

In addition, the follow-up periods included varied from three months (Bennett Cattaneo & Goodman, 2003) to 18 months (Bell et al., 2008; Bennett Cattaneo et al., 2007). However, in the majority of studies, the time of reference that survivors were asked to consider in their predictions, was not matched by the point at which follow-up data collection occurred. For example, Weisz et al. (2000) asked survivors to predict their victimization during the following calendar year, their retrospective reports of abuse were taken only four months later. This was similarly the case in the Bennett Cattaneo and Goodman (2003) study in which participants were asked about the potential for abuse during four month period, but were then contacted from three months after their assessment. Heckert and Gondolf (2004) report that women were assessed at intake and asked to predict continued abuse during the next three months, but then report on outcome data obtained at a minimum of nine months post-intake. Bell et al. (2008) and Bennett Cattaneo et al. (2007) asked survivors to estimate the likelihood of abuse during a 12-month period, but did not follow up until 18 months later.

This varied practice may have impacted on the number of misclassified cases identified, particularly in instances where follow up was earlier than victims expected, due to their existing a period of time during which the outcome was not measured (Weisz et al., 2000). Consequently, those who had not experienced abuse by the time of report, might have done so afterwards and within the time period they had been asked to consider.

The samples used in the studies were also ethnically and socioeconomically diverse, which raises questions about the extent to which such features may have influenced the results, and also the generalizability of the findings. Bennett Cattaneo and Goodman (2003; Bennett Cattaneo et al., 2007), for example, acknowledge that their sample was comprised of mainly of low income African-American women who are more likely to report higher levels of traumatization. Moreover, their sample was drawn from a population of survivors whose partners had been arrested for IPV, as was the case in all studies, and who themselves volunteered to participate in the study. A comparison of those who did and did not complete the follow-up interview indicated, that those who did not participate were at higher risk than those that did which is likely to place them above a particular threshold of risk perception (Bennett Cattaneo et al., 2007).

4.3.2. Factors associated with victim accuracy

Three studies have gone beyond simply examining the accuracy of survivors' predictions of future IPV and have examined factors that influence survivors' perceptions and their accuracy (Bell et al., 2008; Bennett Cattaneo et al., 2007; Gondolf & Heckert, 2003). Gondolf and Heckert (2003) interviewed survivors as part of their multi-site treatment evaluation (n = 443). Interviews were conducted at intake and women were asked how likely they felt their partner was to use violence again in the next three months and how safe they were. In addition, 33 and 36 months after intake women were asked an open ended question concerning their rationale for their answers. The potential predictors of women's responses examined included social demographic characteristics, relationship characteristics, previous IPV perpetrator behaviors (violence, substance use, and contact with

Table 4
Summary of empirical studies that have examined the predictive utility of victim risk appraisals.

Authors/date	Sample	IVs	DVs	Main findings
Weisz et al. (2000)	177 female victims	Victim rating of violence likelihood in next year (0–10 scale)	Severe physical violence assessed at 4 month follow up	Survivors prediction of abuse in following year had greatest bivariate association with outcome (r = .42).
	126 (71.6%) European American	Items assimilating Danger Assessment items: Abuse in preceding six months		Survivors predictions added significantly to regression models (F = 20.85, p = <.001) based on multivariate risk
	85 (48%) married and cohabiting with abuser	Alcohol/drug use by abuser during index offense		
	37 (20.9%) married and separated from abuser	Abuser employment status		
	29 (16.4%) not married but cohabiting with abuser	Victim demographics		
	23 (13%) not married or cohabiting with abuser	Perpetrator controlling behaviors		
	20 (11%) had spoken to a shelter about abuse	Number of children		
Bennett Cattaneo and Goodman (2003)	169 women	Victim rating of violence likelihood in next three months (0–10 scale)	Continued abuse assessed at 3 month follow up	Frequency perpetrator drunk, the generality of violence, scores on the CTS sexual coercion subscales, PMWI dominance-isolation subscale scores and victim assessments were all significant predictors of abuse (n = 96).
	151 (91.1%) African American	Perpetrator age		
	32 (18.9%) married to abuser	Months living together		
	79 (46.7%) dating abuser	Socio-economic status		
	42 (24.9%) separated/ divorced/estranged	Frequency perpetrator drunk		
		Frequency perpetrator high		
		Generality of violence		
		CTS subscale scores		
		Psychological Maltreatment of Women Inventory (PMWI, Tolman, 1989) subscale scores		
Heckert and Gondolf (2004)	499–542 women	Victim rating of violence likelihood in next three months (1–5 scale)	Continued abuse and re-assault at 15 month follow up	Individual risk factors in combination with women's appraisals of future violence were best predictors (ROC AUC .83, sensitivity 70%).
		Socio-demographics		Combining women's perceptions with structured risk assessments improved predictive accuracy, but none surpassed this level.
		Perpetrator alcohol and drug use		
		Prior abuse		
		Personality		
		Women's help seeking		
		K-SID[a]		
		SARA[b]		
		DA[c]		
Bennett Cattaneo et al. (2007)	246 women	Victim rating of violence likelihood in next year (1–5)	Physical assault, injury and attempts to kill, collapsed into one reabuse variable at 18 month follow up	66% of victims accurately assessed their risk
	79% African American	Grouped as high or low risk		No systematic bias in estimates
				Victims equally skilled in predicting re-abuse and non-re-abuse
Bell et al. (2008)	244 women	Victim rating of re-abuse (control/ domination and humiliate/degrade) likelihood in next year (1–5)	Psychological abuse at 18 month follow-up	62% of victims accurately assessed their risk
	80% African American	Grouped as high or low risk		No systematic bias in estimates
				Victims equally skilled in predicting re-abuse and non-re-abuse
Snider et al. (2009)	400 recent victims of IPV	DA	Severe or potentially lethal violence, modified CTS	
		Victim rating of being seriously harmed	9 month follow up	

[a] K-SID = Kingston Screening Instrument for Domestic Abuse.
[b] SARA = Spousal Assault Risk Assessment.
[c] DA = Danger Assessment.

social systems due to violence), IPV perpetrator personality and metal health characteristics, and the survivor's prior help-seeking behaviors.

Quantitative analyses identified eight variables that significantly predicted risk perceptions. These were (in order of importance — greatest first) the perpetrator being high or drunk in the last year, voluntary referral status, perpetrator unemployment, the use of controlling behaviors in the last three months, the use of threats, being separated or divorced and prior severe violence. The weakest predictor was the woman being frequently drunk in the past year which was associated with the perception that they were less likely to be a repeat victim.

Bennett Cattaneo et al. (2007) and Bell et al. (2008) each used the same methodology to examine predictors of survivor accuracy in predicting physical reassault and psychological abuse from the same

cohort of survivors. In each instance, based on predictions and retrospective reports of actual victimization during an 18-month period, survivors were categorized as either true or false positives, true or false negatives. Predictions of accuracy were made from a multivariate model which, in each case, included the variables of PTSD symptoms, substance use, recency of the violence, length of relationship, duration of victimization, physical abuse, psychological abuse, stalking, previous partner violence (with previous partners), availability of social support, and use of formal and informal support. In each instance these variables accounted for 39% and 40% of the variance in accuracy respectively.

In the case of physical abuse, at the univariate level, all variables except duration of abuse and relationship length were associated with accuracy. For psychological abuse, in addition to these two variables,

experience of prior partner violence was also not significantly associated with accuracy. At a multivariate level, two factors predicted survivor's accuracy for both physical and psychological abuse: PTSD symptoms and experiences of stalking. In the case of physical abuse, an increase in PTSD symptoms was associated with a greater likelihood of women being false positives — that is, when women were not accurate they were more likely to overestimate their level of risk. However, PTSD scores were not associated with the likelihood of being inaccurate per se. In other words, PTSD scores did not predict being inaccurate; however, when individuals were inaccurate, PTSD scores were associated with women over — rather than underestimating their level of risk. These findings were broadly replicated for psychological abuse. Prior experience of stalking led women to estimate their risk as high, and was associated with them being accurate in their appraisal (true positives). Indeed, they were more likely to be true positives than any other group when predicting physical and psychological abuse.

Two additional factors were identified as uniquely associated with predicting re-assault: survivor substance use, and the severity of physical abuse experienced. Women who reported substance use were more likely to underestimate their risk (false negatives) than any other group. Severity of physical abuse was associated with false negative status — that is, those individuals who perceived themselves as low risk despite increased severe past violence were more likely to have underestimated their risk.

Two further factors were identified as uniquely associated with predicting psychological abuse: the recency of the latest episode of abuse and prior psychological abuse. The more temporally remote the most recent episode of abuse, the less likely victims were to underestimate their risk, indicating that the passage of time did not make victims less accurate in their risk perception. However, it was found that when mistakes were made, the greater the temporal distance the more likely victims were to underestimate their risk of future psychological abuse. In relation to prior psychological abuse, women who had experienced more severe psychological abuse were more likely to estimate their future risk as high, but were equally likely to be correct as incorrect.

4.3.3. Summary

It is evident that victims do often provide clinically relevant and important information. Indeed, even naive ratings of the perceived likelihood of future victimization yield predictive accuracy estimates that are on par with those achieved by many of the existing structured risk assessment instruments. That the accuracy of victim appraisals appears to be systematically influenced by a range of factors including past experience of stalking and PTSD symptomology provides a basis for contextualizing and expanding the remit of victim assessments when incorporating such appraisals into risk assessments. There are clear implications for risk assessment practice.

5. Discussion

Since 2000, when Dutton and Kropp published their review of the two main IPV specific risk assessment tools, it is clear that there has been some progress made with regard to the development and validation of risk assessment instruments. However, progress in this area still lags far behind that documented for sexual offenders and generally violent offenders for which in 2007, Hanson, Helmus and Bourgon identified at least 79 studies for sexual offenders and 88 studies (post-1980) for violent recidivism. It is evident from the literature reviewed here, that the existing risk assessment instruments show a modest level of predictive accuracy. Moreover, the evidence suggests that actuarial instruments outperform those based on structured professional judgment approaches, and that particularly problematic is the reliability of the aspects of structured professional judgment-derived instruments that typify clinical judgment (e.g., the identification of critical items; Kropp & Hart, 2000). Consequently, it is also clear that there is still room

for improvement with regard to the reliability and validity of IPV risk assessment instruments. Indeed, evidence exists to question the utility of developing IPV specific risk assessment instruments. For example, despite using a contrived psychiatric sample, Hilton, Harris, and Rice (2001) found that the VRAG outperformed the best estimates of the predictive validity of the SARA identified in this review (ROC = .75). In a recent meta-analysis of 33 studies of the predictive accuracy of a range of IPV risk prediction methods, Hanson et al. (2007) found evidence that structured risk assessments for general violence marginally outperformed IPV specific risk assessment tools ($d = .54$ vs. $d = .40$), and that women's appraisals of risk performed almost as well as IPV specific risk assessment tools ($d = .40$ vs. $d = .36$ respectfully), although these differences were not significant. Moreover, both structured professional judgment approaches and women's appraisals performed equally well ($d = .36$).

5.1. Implications for research

The literature reviewed has highlighted a range of areas which require additional research attention. First, more research is required to validate the instruments described herein across a wider range of socioeconomic and cultural groups from a broader range of countries. For example, although the SARA is widely used in England and Wales in order to treatment match in prison, and identify appropriate levels of multi-agency supervision in the community (Bowen, 2011), to date, there is no published validation study of the SARA on a British population. In addition, despite a recent acknowledgement of partner violence towards men by female intimate partners, and the use of violence within same sex relationships, there exists a dearth of literature that has examined risk factors for these forms of research, and only one study that has attempted to modify an existing risk assessment for use to assess risk of violence in lesbian relationships (Glass et al., 2008). This lag in development is understandable given the reality that the majority of cases that come into contact with criminal justice agencies are those of violence against women in heterosexual relationships. However, given the pro-arrest policies which have resulted in an increased number of women being arrested within the context of IPV, it would be timely to pursue this avenue of research further.

The research conducted so far has provided useful information concerning the factors that might influence the accuracy of survivor's predictions concerning their future victimization. However, more attention needs to be directed to understanding the process through which victims reaches these decisions, and the extent to which these processes differ as a function of the clinical characteristics identified as relevant to predictive accuracy. Gondolf and Heckert (2003) asked survivors how they reached their decisions regarding their level of risk. It was found that women's perceptions of risk were based on their appraisals of their partner's behavior and the current relationship, in other words, women seemed to base their judgment on the presence of some of the standard risk factors which are integrated within risk assessment instruments. However, that women's predictions can contribute meaningfully to the predictive accuracy of these tools indicates that there must be additional contextual factors which influence their appraisals. These need to be examined in more depth to determine whether existing instruments should be adapted to incorporate them to increase their overall predictive accuracy. Perhaps even more importantly, it would be useful to determine the factors that predict women's accuracy in identifying the cessation of abuse, and, in particular, potential protective factors which are notoriously absent from standard instruments (Rogers, 2000). It might be that the inclusion of protective factors may increase the predictive validity of risk assessment tools.

More research is required to determine how women's perceptions of risk are formulated and how such perceptions affect their behavior with regards to help seeking and safety (Heckert & Gondolf, 2004). For example, is it the case that in circumstances of perceived high risk,

victims alter their behavior to reduce the likelihood of future victimization? Perhaps more importantly, under such circumstances are there factors that reduce the likelihood of women seeking safety?

What must also be noted, however, is the fact that no form of prediction is perfect. In the case of women's perceptions, approximately one third of women were found to be incorrect in their estimates of risk (Bell et al., 2008; Bennett Cattaneo et al., 2007). Consequently, more research is required to understand the factors that lead to inaccurate perceptions of risk, and how practitioners can help survivors to appreciate the level of risk that they face.

5.2. Implications for practice

The reliability and validity data presented in this review are insufficient to promote the use of one type of measure over and above another due to the limited amount of validation data available. Consequently, practitioners engaged in IPV risk assessment need to be guided in their selection of approach by the requirements placed upon them and the context and purpose of the risk assessment. However, Heckert and Gondolf (2004) note that simulated risk assessments (i.e., those that are not conducted as designed [e.g., based on file information only], or derived from an existing risk factor battery) should be used with caution due to their reduced levels of predictive accuracy.

Given the comparative recency of the development of the actuarial measures, it would be prudent to adopt these measures as and when sufficient relevant and appropriate validation data exist so that more valid conclusions about an individual's potential for re-offending can be drawn from an appropriate normative sample. Indeed, as with the use of any actuarial instrument, making individual assessments based on group level aggregate reoffending data requires that the individual share as many relevant characteristics with the normative group as possible. Otherwise any statistical predictions are likely to be of limited validity (Craig & Beech, 2010).

The evidence concerning the validity of victims' appraisals of their own risk provides further support for best practice assessment guidelines which emphasize obtaining partner reports. However, rather than simply using partner reports to validate perpetrator accounts the data reviewed here suggest that particular attention should be paid to how safe a survivor feels and the extent to which she believes that she is likely to be victimized in the future. The studies examined indicate that any claims by survivors that they feel unsafe or believe that they are at risk should be taken seriously (Weisz et al., 2000). This should be the case even if risk factors do not indicate that a risk is present (Gondolf & Heckert, 2003). It is likely that victims are sensitive to more idiosyncratic features of their relationship and their abuser than can be identified through standard risk assessments alone, and consequently this explains why victim appraisals add to the predictive accuracy of standard risk assessment tools (Bennett Cattaneo et al., 2007). To that end, the data examined also have implications for the appropriate use of risk assessment instruments (Weisz et al., 2000). Ultimately, in order to ensure that such practice is conducted, multi-agency working is required with clear information sharing protocols so that individuals working with perpetrators can obtain partner information. Although such practice is widely endorsed as best practice, these data reinforce the potential benefits that might result.

5.3. Conclusion

Although there has been recent growth in the development of risk assessment tools for IPV, the field currently remains under-developed. The existing validation literature is limited and considerable work is required in order to provide practitioners with the ability to make evidence based decisions regarding their choice and use of these instruments. In addition, current risk assessment practice would benefit from greater emphasis on integrating the unique insights provided by victims into the risk appraisal process, although more

research is required in order to understand which factors should be most clearly attended to. What is evident, however, is that victims have access to a range of information relevant to their personal risk and that practitioners need to pay attention to claims of increased risk if we are to go some way to reducing IPV re-offending.

References

Ægisdottir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counselling Psychologist, 34*, 341−382.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed). Cincinatti, OH: Anderson Publishing Company.

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.

Bell, M. E., Bennett Cattaneo, L., 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*, 69−80.

Bennett Cattaneo, L., 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*, 429−440.

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

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

Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *Journal of Psychiatry and Law, 21*, 35−76.

Bowen, E (2011). *The rehabilitation of partner-violent men.* Chichester: Wiley-Blackwell.

Browne, A. (1987). *Battered women who kill.* New York: Free Press.

Campbell, J. C. (1995). Prediction of homicide of and by battered women. In J. Campbell (Ed.), *Assessing Dangerousness: Violence by Sexual Offenders, Batterers, and Child Abusers* (pp. 96−113). Thousand Oaks, CA: Sage.

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*, 653−673.

Coontz, P. D., Lidz, C. W., & Mulvey, E. P. (1994). Gender and the assessment of dangerousness in the psychiatric emergency room. *International Journal of Law and Psychiatry, 17*, 369−376.

Craig, L. A., & Beech, A. R. (2010). Towards a guide to best practice in conducting actuarial risk assessments with sex offenders. *Aggression and Violent Behavior, 15*, 278−293.

Douglas, K. C., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology, 4*, 149−184.

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

Doyle, M., & Dolan, M. (2008). Understanding and managing risk. In K. Soothill, P. Rogers, & M. Dolan (Eds.), *Handbook of Forensic Mental Health*. Uffculme, Devon, UK: Willan Publishing.

Dutton, M. A. (1996). Battered women's strategic response to violence: the role of context. In J. Edleson, & 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.

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

Elbogen, E. B. (2002). The process of violence risk assessment: A review of descriptive research. *Aggression and Violent Behavior, 7*, 591−604.

Garcia-Moreno, C., Jansen, H. A. F. M., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: Findings from the WHO multi-country study on women's health and domestic violence. *Lancet, 369*, 1260−1269.

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*, 1021−1027.

Gondolf, E. W., & Heckert, D. A. (2003). Determinants of women's perceptions of risk in battering relationships. *Violence and Victims, 18*, 371−386.

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*, 63−74.

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

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

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

Grubin, D. (1997). Inferring predictors of risk: sex offenders. *International Review of Psychiatry, 9*, 225−232.

Hanson, R. K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis.* Ontario, CA: Public Safety Canada.

Hare, R. (1991). *The Hare Psychopathy Checklist Revised.* New York: Multi-Health Systems.

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*, *20*, 315–335.

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

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*, 778–800.

Hilton, N. Z., & Harris, G. T. (2007). Assessing risk of partner violence. In J. Campbell (Ed.), *Assessing Dangerousness: Violence by Batterers and Child Abusers* (2nd ed.). New York, NY: Springer.

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

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2010). *Risk assessment for domestically violent men.* Washington, DC: American Psychological Association.

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.

Home Office (2005). *Domestic violence: A national report.* London: Author.

Kropp, P. R. (2004). Some questions regarding spousal assault risk assessment. *Violence Against Women*, *10*, 676–697.

Kropp, P. R. (2007). Spousal assaulters. In C. D. Webster, & S. J. Hucker (Eds.), *Violence Risk Assessment and Management* (pp. 123–132). Chichester: Wiley.

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*, 101–118.

Kropp, P. R., & Hart, S. D. (2004). *The development of the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER): A tool for criminal justice professionals.* Otawa: Department of Justice Canada.

Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER): User manual.* Vancouver, BC: ProActive Resolutions Inc.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Manual for the Spousal Assault Risk Assessment Guide* (3rd ed.). Toronto, Canada: Multi-Health Systems.

Kury, H., Obergfell-Fuchs, J., & Woessner, G. (2004). The extent of family violence in Europe: A comparison of national surveys. *Violence Against Women*, *10*, 749–769.

Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. *JAMA, the Journal of the American Medical Association*, *269*, 1007–1011.

McNeil, D. E., & Binder, R. L. (1995). Correlates of accuracy in assessment of psychiatric inpatients risk of violence. *The American Journal of Psychiatry*, *148*, 1317–1321.

McNeil, D. E., Sanburg, D. A., & Binder, R. L. (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric patient's potential for violence. *Law and Human Behavior*, *22*, 655–667.

Monahan, J. (1981). *The clinical prediction of violent behaviour.* Rockville, MD: National Institute of Mental Health.

Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behavior*, *23*, 107–120.

Moore, B. (1996). *Risk assessment: A practitioner's guide to predicting harmful behavior.* London: Whiting & Birch, Ltd.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783–792.

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

Povey, D., Coleman, K., Kaiza, P., & Roe, S. (2009). *Homicides, firearms and intimate violence 2007/08: Supplementary volume 2 to Crime in England and Wales 2007/08.* Home Office Statistical Bulletin 02/09 (3rd ed). London: Home Office.

Quinsey, V. L. (1995). The prediction and explanation of criminal violence. *International Journal of Law and Psychiatry*, *18*, 117–127.

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

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

Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*, *24*, 595–605.

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

Straus, M. A. (1979). Measuring intra-family conflict and violence: The conflict tactics scale. *Journal of Marriage and the Family*, *75*–88.

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

Walby, S., & Allen, J. (2004). *Domestic violence, sexual assault and stalking: Findings from the British Crime Survey.* London: Home Office.

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

Whittemore, K. E., & Kropp, P. R. (2002). Spousal assault risk assessment: A guide for clinicians. *Journal of Forensic Psychology Practice*, *2*, 53–64.

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*, 437–455.