IHE Report

Spousal Violence Against Women: Preventing Recurrence

January 2008

IHE
INSTITUTE OF HEALTH ECONOMICS
ALBERTA CANADA

Institute of Health Economics

The Institute of Health Economics (IHE) is an independent, not-for-profit organization that performs research in health economics and synthesizes evidence in health technology assessment to assist health policy making and best medical practices.

IHE Board of Directors

RISK ASSESSMENT TOOLS FOR PREDICTING RECIDIVISM OF SPOUSAL VIOLENCE

Prepared by:

Bing Guo
Christa Harstall

ACKNOWLEDGEMENTS

The Institute of Health Economics is most grateful to the following individuals for review and provision of information and comments on the draft report The views expressed in the final report are those of the Institute.

— Dr. Donald G. Dutton, University of British Columbia, Vancouver, Canada
— Dr. Janice Roehl, Justice Research Center, Pacific Grove, USA
— Mr. Gary Gibbens, YWCA Sheriff King Program, Calgary, Canada

The authors also wish to thank the following individuals for providing information regarding the local context or published studies:

— Dr. Jacquelyn C. Campbell, Johns Hopkins University School of Nursing, Baltimore, USA
— Dr. Sue Ludwig, Alberta Mental Health Board, Edmonton, Canada

Information Service Support

The literature search for the review was undertaken by Ms. Liz Dennett, Information Specialist, Institute of Health Economics and University of Alberta, Edmonton, Canada.

Competing interest

Only the referees who were paid honoraria were asked to sign a declaration of competing interest. All others who provided general information were exempt.

Competing interest is considered to be financial interest or non-financial interest, either direct or indirect, that would affect the research contained in this report or create a situation in which a person's judgment could be unduly influenced by a secondary interest such as personal advancement.

On the basis of the statement above, Dr. Dutton and Dr. Roehl declared no competing interest.

Production of this document has been made possible by a financial contribution from Alberta Health and Wellness. The views expressed herein do not necessarily represent the official policy of Alberta Health and Wellness.

EXECUTIVE SUMMARY

Background

Male-to-female spousal violence is a significant public health problem because of the associated physical, psychological, and financial costs to victims, their families, communities, and society in general. Worldwide 10% to 69% of women reported being physically assaulted by an intimate male partner at some point in their lives. The 2004 General Social Survey indicated that approximately 7% of Canadian women 15 years of age and over in a current, previous, or common-law union experienced spousal violence in the previous five years.

To prevent future serious or lethal assaults and to help allocate already scarce resources, risk assessment instruments have been developed in an attempt to identify offenders who are at high risk of reoffending. However, little is known about the predictive validity of these instruments.

Objective

To assess the research evidence on the inter-rater reliability and predictive validity of various risk assessment instruments in predicting male-to-female spousal violence recidivism and lethality in those males who had contact with the police system.

Methodology

All original published systematic reviews or primary studies were identified by systematically searching PubMed, CRD databases (NHS EED, HTA, DARE), EMBASE, Family & Society Studies Worldwide, Sociological Abstracts, Social Services Abstracts, Social Sciences Abstracts, Academic Search Premier, Web of Science, PsycINFO, and ABI/Inform from January 1995 to May 2007. Relevant library collections and websites of health technology assessment-related agency resources were also searched.

Results

No systematic review on this topic was located through a comprehensive literature search. Eight primary studies were found that evaluated the predictive validity of several currently used instruments, including the Ontario Domestic Assault Risk Assessment (ODARA), the Spousal Assault Risk Assessment (SARA), the Danger Assessment (DA), the Domestic Violence Screening Instrument (DVSI), the Violence Risk Appraisal Guide (VRAG), and the Level of Service Inventory-Revised (LSI-R). The ODARA was evaluated in only one Canadian study, whereas the SARA was evaluated in five studies.

The characteristics of the population under study–male offenders–varied considerably across the included studies. Some male offenders were arrested, on probation, or in a maximum-security psychiatric facility, whereas others were referred to attend batterer treatment programs that were provided in Canada or the United States.

Inter-rater reliability was tested for the SARA, the ODARA, and the VRAG, but not for the DA and other instruments. Limited research indicated good inter-rater reliability for the SARA, the ODARA, and the VRAG.

In terms of predictive validity, five studies compared the predictive validity of one instrument with other instruments, whereas three studies reported the predictive validity of only one instrument. None of the studies reported any lethal assault during the follow up period. The reported area under the receiver operating characteristic curve (ROC AUC, a measure of predictive validity) was less than 0.80 (ranged from 0.59 to 0.77) for all instruments under evaluation, suggesting only marginal to moderate improvement over chance in predicting non-lethal recidivism.

Only one study that was conducted in Canada compared predictive validity of the ODARA, the SARA, and the DA. On the basis of the ROC analysis, the predictive validity of the ODARA appeared to be lower in the cross-validation sample than that in the construction sample. In the construction sample, all three instruments yielded statistically significant predictive validity, but the ODARA predicted recidivism statistically significantly better than did the SARA or the DA. In the cross-validation sample, the ODARA significantly predicted recidivism, whereas the DA and the SARA did not. These findings need to be interpreted cautiously. First, instructions for the SARA were not completely followed; that is, no interview was performed, and no clinical judgment was made. Second, no lethal assault occurred during the follow up in this study; therefore, it could not be concluded that the ODARA was superior to the DA in predicting lethal assault.

Conclusions

All instruments under evaluation demonstrated improvement over chance in predicting spousal violence recidivism; however, no conclusion could be made regarding the superiority of one tool over another at this time. Although lethal assault is of the greatest concern, the included studies failed to provide any information on how well these instruments (even the DA) predict lethal assault. Current research evidence on the predictive validity of these instruments has been exclusively based on measuring non-lethal reassault.

The decision on selecting an appropriate risk assessment instrument needs to take into account factors such as the available research evidence, the population under assessment, the intended users of the instrument, and the

purpose for conducting an assessment. The DA may be most appropriately used by clinicians, victim advocates, or social workers in women's shelters, hospitals, and women's treatment programs. The ODARA, on the other hand, may be more appropriately applied by law enforcement personnel. The SARA may not be an optimal instrument for use by law enforcement personnel but it provides an immediate opportunity for the development of a comprehensive treatment plan for the assaulter.

Because the research evidence available on the predictive validity of risk assessment instruments is very limited, it would be inappropriate to make any decision of an individual's risk of recidivism based solely on the scores of the instruments being used. Information from all other sources, such as women's perceptions of risk, and records in the justice system, should be gathered to make more accurate predictions.

The Alberta Mental Health Board is in a unique situation where both the SARA and the ODARA are being applied to the target population of interest and within the local context. Data gathered during the implementation of the Provincial Family Violence Treatment Program using these risk assessment tools will provide valuable information regarding their predictive validity and the initiation of a classification system of recidivism severity. This implementation and evaluation process will allow a direct comparison of the predictive validity of the SARA, the ODARA, or a combination of both, in the subgroup of male abusers who were referred to the treatment program, to see whether the combination of the two tools would improve predictive validity. Findings from this process would be helpful to determine whether a composite tool that fits the Alberta context could be developed and used province-wide.

ABBREVIATIONS

AMHB – Alberta Mental Health Board

AUC – Area under the curve

CTS2 – Conflict Tactics Scales (Revised)

DA – Danger Assessment

DV – Domestic violence

DVSI – Domestic Violence Screening Instrument

DVSR – Domestic Violence Supplementary Report

H-10 – Historical part of HCR-20

HCR-20 – Historical Clinical Risk-20

ICC – Intra-class correlation coefficient

K-SID – Kingston Screening Instrument for Domestic Violence

LSI-R – Level of Service Inventory - Revised

ODARA – Ontario Domestic Assault Risk Assessment

PCL-R – Psychopathy Checklist-Revised

ROC curve – Receiver operating characteristic curve

SARA – Spousal Assault Risk Assessment

SEM – Standard error of measurement

VRAG – Violence Risk Appraisal Guide

Table of Contents

List of Tables

OBJECTIVE AND SCOPE

This report was prepared in response to a request from the Alberta Mental Health Board to conduct a critical appraisal of the literature on the validity and reliability of various risk assessment instruments for predicting recidivism and lethality for male spousal violence offenders.

The objectives of this report are:

1. to identify studies that evaluated various risk assessment instruments, and to assess the research evidence on their reliability and validity, with an emphasis on inter-rater reliability and predictive validity, in predicting male-to-female spousal violence recidivism and lethality in those males who had contact with the police system; and
2. to identify and assess studies that compared the predictive validity of two or more risk assessment instruments.

Studies that focused on female-to-male spousal violence, child abuse, or elderly abuse; on predicting onset rather than recidivism of spousal violence; and on the victim's perception of risk were beyond the scope of this report.

BACKGROUND

Definition of spousal violence

A number of different terms, such as family violence, domestic violence, spousal violence, women abuse, wife abuse, wife assault, spouse abuse, and intimate partner violence, are used interchangeably to describe the abuse of a woman by her partner, family members, caregivers, or others with whom she has intimate, familial, or romantic relationships.[1,2] All of these terms reflect a common understanding of violence as a manifestation of power, control, and domination that is expressed through a range of ongoing and escalating violent behaviours.[2] The term 'spousal violence' was chosen for this report for two reasons: (1) This term excludes other types of domestic or family violence, such as child abuse or elderly abuse, and (2) this is the term that is used by Statistics Canada in its annual report on family violence. In this report, spousal violence mainly refers to male-to-female violence, which is considered the most serious form of spousal violence because of its prevalence, repetitive nature, and high risk of morbidity and mortality.[3]

Spousal violence can occur in many forms, such as physical, emotional/psychological, sexual, verbal, environmental, social, financial, or religious/spiritual abuse (Table 1). Physical abuse usually occurs in conjunction with sexual and psychological abuse.[4]

Table 1: Various forms of spousal violence

Spousal violence	Definition
Physical abuse	Includes slapping, punching, kicking, biting, shoving, choking, or using a weapon to threaten or injure a woman. It can include any unwanted physical contact or physical neglect, and it may result in death.
Emotional/ psychological abuse	Includes various forms of intimidation, harassment, excessive jealousy, control, isolation, and threats.
Sexual abuse	Includes any forced sexual activity or infecting a women with a sexually transmitted disease by engaging in unsafe sexual practices.
Verbal abuse	Includes constant criticism, blaming, false accusations, name calling, and threats of violence toward a woman or people or things she cares about.
Environmental abuse	Includes making a woman feel afraid in her home or environment by destroying property and possessions as a form of intimidation.
Social abuse	Includes isolating the woman from her friends and family.
Financial abuse	Includes preventing a woman from having financial independence, economically exploiting her, or preventing her from having any control over the family's money and expenditure decisions.
Religious/ spiritual abuse	Involves ridiculing a woman's beliefs, using her beliefs to manipulate her, or denying her or her children's involvement in her spiritual or religious practice.

Adapted from Jamieson et al. 1999[1]

Epidemiology

Spousal violence has long been recognized as a significant public health problem because of the associated physical, psychological, and financial costs to victims, their families, and society in general.[5] Worldwide between 10% and 69% of women reported being physically assaulted by an intimate male partner at some point in their lives.[5]

According to the 2004 General Social Survey, approximately 7% of Canadian women, representing an estimated 653,000 women, 15 years of age and over, in a current, previous, or common-law union, experienced spousal violence in the previous five years.[8] The Canada Violence Against Women Survey found that more than half (56%) of the women who had experienced spousal violence in the year prior to the survey were 18 to 34 years of age, a range that coincides with the main childbearing years.[1] The annual prevalence of spousal violence during pregnancy was reported as between 6% and 8%.[2] Aboriginal people (including Indians, Métis, and Inuit) were three times more likely to be victims of spousal violence than were those who were non-Aboriginal.[8]

In 2004, 5,520 spousal violence incidents were reported to police in Alberta, resulting in 3,865 charges.[9] Statistics Canada reported that women in Alberta experience the highest rate (10%) of domestic violence in Canada.[9] In 2003/04, 5,929 women and 5,558 dependent children were sheltered in Alberta.[9]

Spousal violence is repetitive and can sometimes be fatal. According to Statistics Canada, half of those who self-reported spousal violence stated that the violence occurred on more than one occasion, and women are much more likely than men to report that they were the targets of more than 10 violent incidents at the hands of their current or previous partners.[8] Between 1994 and 2003, spousal homicide represented approximately one in five solved homicides (homicides in which at least one accused was identified by police) in Canada and almost half (47%) of all solved family homicides.[8] In 2003, 78 persons were killed by their spouse, of whom 64 were female victims.[8] Between 1994 and 2003, spousal homicide rates were lowest in the Atlantic provinces (3.47 per million spouses). Rates of spousal homicide among the four Western provinces were higher than all other regions, with Saskatchewan reporting the highest rate of spousal homicide (7.6 per million spouses).[8] Aboriginal women in Canada are eight times more likely than non-Aboriginal women to be killed by their partners.[10]

According to the 1995-96 US National Violence Against Women Survey,[11] 25% of the 8,000 surveyed women were abused by their intimate partners at some time in their lives, whereas 1.5% of all surveyed women were abused by their partners in the previous 12 months. From these data, it was estimated that approximately 1.5 million women were abused by an intimate partner annually in the United States. Forty-one percent of physically assaulted women had injuries.[11] Intimate partner homicide is the largest category of murders of women, accounting for approximately 30% to 40% of murders according to the official counts based on the Supplemental Homicide Reports.[12] In 1997, 1,174 women were killed by their intimate partners.[13]

Violence against women is almost universally underreported because of the sensitivity of the subject.[2] The reported rates therefore should be considered as representing the minimum levels of violence that occur.[2]

Health consequences

Spousal violence causes both physical and mental health problems, such as physical injury, chronic pain syndrome, gastrointestinal disorders, reproductive health consequences, depression, suicide attempts, and psychosomatic disorders.[5] Abuse during pregnancy may have consequences beyond injury, including women's problems with anxiety, depression, anemia, infections, and substance use. In addition, abused pregnant women may experience pregnancy related problems such as inadequate prenatal care and low pregnancy weight gain.[14] The US Centers for Disease Control and Prevention reported that the healthcare costs of intimate partner rape, physical assault, and stalking exceed $5.8 billion each year, with $4.1 billion going toward medical and mental healthcare services.[15]

Risk assessment for spousal violence recidivism

The severe consequences of repeated spousal violence, especially lethal assault, necessitate the development of ways to identify male abusers at high risk of recidivism so that actions can be taken to protect victims. On the other hand, the increased use of the criminal justice system for acts of spousal violence has necessitated a sort of triage in case processing by law enforcement, prosecutors, and the courts.[12] In the United States, proactive arrest policies emerging in the past two decades have produced an overwhelming number of spousal assaulters entering the criminal and civil justice systems.[6] This increase has strained already inadequate resources for managing offenders and protecting victims of spousal violence.[6] Determining the seriousness of a particular case and risk of escalation is necessary not only for the obvious reason of allocating resources, but also to tailor the response to the level of dangerousness. The response needs to be appropriate to avoid violating the civil rights of offenders and unnecessarily disrupting the lives of victims and their children.[12] In response to this need, mechanisms such as checklists, clinical interviews or practitioners' intuitions, and formal risk assessment instruments have been adopted to identify high-risk offenders.[12]

Risk assessment can serve as the cornerstone for offender management in a variety of contexts, such as evaluating an offender's suitability for a particular program, determining whether to warn a spouse regarding imminent danger, choosing the appropriate sentence or sanction, evaluating the needed level of care, ascertaining whether discharge from a psychiatric or mandated program is warranted, determining parental fitness for retaining custody of children, or making a judgment about needed services at the request of social service agencies, law enforcement workers, or the courts.[16] The following benefits of using risk assessment have been identified:[17]

- to assist women and spousal violence workers to develop more realistic safety plans;
- to assist in the development and delivery of appropriate services within perpetrators' treatment programs;
- to help criminal justice systems to identify offenders who need closer supervision;
- to educate service providers about spousal violence; and
- to provide a shared language about 'risk' for service providers who may come from a range of different agencies.

RISK ASSESSMENT INSTRUMENTS

Approaches to risk assessment

A major problem in conducting risk assessments of spousal violence recidivism is the lack of a systematic, standardized, clinically useful, and empirically based framework for collecting, weighting, and reporting background data and professional judgments.[3] Only in the last 10 to 15 years have predictive items for spousal assault recidivism been empirically identified.[18]

Risk assessment instruments have been developed to predict the likelihood of future recidivism. These instruments generally reflect one of three approaches to risk assessment: (1) unstructured clinical risk assessment, (2) actuarial risk assessment, and (3) structured clinical judgment.[19]

Unstructured clinical risk assessment

Unstructured clinical assessment is probably still the most widely used approach.[20] This method involves no constraints or guidelines for the assessor. Decisions are based on professional discretion and are usually justified according to the qualifications and experience of the professional.[20]

This approach offers the advantage of flexibility and focuses on violence prevention.[19] It allows for an idiographic, or individual-centered, analysis of the offender's behaviour and a context-specific tailoring of risk and violence management strategies.

This approach, however, has been widely criticized for its low inter-rater reliability, low validity, failure to specify the decision-making process, and inferior predictive validity compared with actuarial predictions.[19,20] Because this approach relies heavily on professional discretion, it is vulnerable to missing important factors that may require intervention. Recommendations for management strategies might be based more on the training, preferences, and biases of the assessor rather than on well-reasoned consideration of dynamic and crime-relevant risk factors and intervention strategies that are either empirically valid or well accepted in the field.[20]

Actuarial risk assessment

An actuarial approach can be defined as an instrument that provides weightings and empirically based scores that relate predictor variables selected from measured associations to criterion variables in representative samples.[12] The actuarial assessment is designed to predict specific behaviours within a specific time frame. The stated goal of the actuarial assessment is to predict violence in

a relative sense, by comparing an individual to a norm-based reference group, and in an absolute sense, by providing a precise, probabilistic estimate of the likelihood of future violence.[20] The actuarial approach allows assessors to make decisions based on data that can be coded in a predetermined manner. Decisions are made according to rules and focus on relatively small numbers of risk factors that are known, or are thought to predict violence across settings and individuals.[19]

The key strength of this approach is that it improves upon the poor reliability and validity of unstructured clinical assessments.[20] However, this approach tends to ignore individual variations in risk, is overly focused on relatively static variables, fails to prioritize clinically relevant variables, and minimizes the role of professional judgment.[19] Practitioners resist using approaches that eliminate professional discretion because they see their role as preventing violence rather than predicting it. To apply the actuarial approach properly, the assessor is forced to consider a fixed set of factors and cannot consider unique, unusual, or context-specific variables that might require intervention.[20]

In Canada, this approach has been used by the Ontario Provincial Police for the development of the Ontario Domestic Assault Risk Assessment (ODARA).[20]

Structured clinical (professional) judgment

Structured professional judgment attempts to bridge the gap between actuarial and unstructured clinical approaches to risk assessment.[20] The structured clinical instruments promote systematic data collection based on sound scientific knowledge, yet allow flexibility in the assessment process. Unlike strict actuarial measures, they encourage clinicians to use professional discretion.[19]

The term 'professional' is used to allow for the reality that there are non-clinical professionals (e.g. police officers, probation officers, victim services personnel) who are often required to conduct risk assessments.[20] This approach has also been termed the 'guided clinical approach.' The assessor must conduct the assessment according to guidelines that reflect current theoretical, professional, and empirical knowledge about violence.[20] Such guidelines provide the minimum set of risk factors that should be considered in every case. The guidelines will also typically include recommendations for information gathering (e.g. the use of multiple sources and multiple methods), communicating opinions, and implementing prevention strategies.[20]

Structured professional judgment does not impose any restrictions for the inclusion, weighting, or combining of risk factors. Items in such instruments are primarily identified from the literature on characteristics of male abusers, predictors of violent crime, and clinical experience, but are not directly derived

from follow-up studies.[18] An advantage of this approach is that a list of items and scoring criteria can be easily generated, and simply computing a total of item scores have a good (but not certain) chance of yielding a correlation with spousal violence recidivism. Because the items are selected from the literature, generalizability may be less problematic than it is for an instrument constructed using risk factors from one particular population.[18] However, subsequent empirical validations are required for data on reliability, accuracy, and population norms.[18] Drawbacks associated with this approach include the need for clinical training and the lengthy time for conducting risk assessment.

In Canada, this approach was used by the British Columbia Institute Against Family Violence in the development of the Spousal Assault Risk Assessment (SARA) Guide.[20]

Components of the instruments

There are two types of risk factors: static or dynamic. Static risk factors refer to demographic or criminological variables, which at any given moment are fixed or determined beforehand (by prior events). Examples include gender, age at first court appearance, or having a parent with a criminal record. In contrast, dynamic risk factors fluctuate over time and reflect internal states or temporary circumstances of the individual, such as beliefs and cognitions, everyday associates, and feelings of hostility. Such factors vary in their speed of change and are sometimes characterized as 'stable' or 'acute'.[21] Risk assessment instruments usually contain both types of risk factors, with emphasis more or less on one type.

Measuring the accuracy of risk assessment instruments

Testing of reliability and validity for an instrument is of great importance in predictive research. There are, however, some challenges around validation of the risk assessment instruments,[22] many of which have yet to be appropriately validated.

Reliability

Reliability refers to the precision and consistency of a measurement, or the extent to which a test score consists of the true score and not measurement error.[16] Reliability is usually determined by the extent to which a score can be replicated in identical or equivalent testing situations.[23] Test-retest reliability, inter-rater reliability, internal consistency reliability, alternate form reliability, and standard error of measurement are some basic ways to evaluate the reliability of a given instrument (Table 2).

Table 2: Types of reliability

Reliability	Definition
Test-retest reliability	The degree to which a score remains stable from Time 1 to Time 2.
Inter-rater reliability	The degree to which multiple observers agree when assigning scores using the same instrument on the same participant within the same time frame (or at the exact same time).
Internal consistency reliability	The extent to which a set of items measure various aspects of the same characteristic and nothing else.
Alternate forms reliability	Involves the preparing of an equivalent form of the test with non-overlapping items and calculating the degree to which the two versions are correlated.
Standard error of measurement (SEM)	An estimate of the standard deviation of error scores of an instrument. The examination of SEM allows for a direct assessment of the degree of error operating in repeated measurements using the same instrument.

Adapted from Rathus & Feindler 2003,[16] Portney & Watkins 2000[24]

Analytic methods for measuring inter-rater reliability

For a risk assessment instrument, inter-rater reliability is much more important than other types of reliability such as internal consistency or test-retest reliability, because a risk assessment instrument is used for forecasting future behaviour rather than measuring the strength of a psychological disposition. To conduct a risk assessment is of little value if raters cannot agree on the presence of individual risk factors or the implications that can be drawn from them.[7]

Inter-rater reliability is usually measured in one of four ways.[16] For dimensional or frequency data (e.g. total scores on self-report measures), inter-rater reliability is normally calculated in terms of intra-class correlation coefficients (ICCs) (range from 0 to 1.0) or simple Pearson correlations between rater's scores. Intra-class correlations are the preferred method because they take concordance of ranks and levels of ratings into account rather than being calculated solely on the basis of concordance of ranks.[16] ICC values are generally considered to be good with scores greater than 0.50 to 0.60.[7]

For categorical data (e.g. types of communication observed in an observational study), inter-rater reliability is normally calculated with the Cohen's kappa statistic (range from 0 to 1.0) or by percentage of agreement (number of agreements divided by the number of agreements plus disagreements). Kappa is generally superior to simple percentage of agreement because its calculation takes chance agreement into account.[16]

Validity

Validity refers to the extent to which an instrument measures what it attempts to measure.[16] Common types of validity are presented in Table 3.

Table 3: Types of validity

Validity	Definition
Face validity	Indicates that an instrument appears to test what it is supposed to test. The weakest form of measurement validity.
Content validity	Indicates that the items that make up an instrument adequately sample the universe of content that defines the variable being measured. Most useful with questionnaires and inventories.
Criterion-related validity	Indicates that the outcomes of one instrument, the target test, can be used as a substitute measure for an established gold standard criterion test. Can be tested as concurrent or predictive validity.
Concurrent validity	Establishes validity when two measures are taken at relatively the same time. Most often used when the target test is considered more efficient than the gold standard and, therefore, can be used instead of the gold standard.
Predictive validity	Establishes that the outcome of the target test can be used to predict a future criterion score or outcome.
Construct validity	Establishes the ability of an instrument to measure an abstract construct and the degree to which the instrument reflects the theoretical components of the construct.

Adapted from Portney & Watkins 2000[24]

Analytic methods for predictive validity

The principal purpose of risk assessments is to estimate the likelihood of overt behaviour, such as interpersonal violence; thus item selection by factor analysis (defined as a set of statistical methods for analyzing the correlations among several variables to estimate the number of fundamental dimensions that underlie the observed data and to describe and measure those dimensions),[25] and construct validity are central to psychological test construction, but less central to formal risk assessment construction. The primary test of risk assessment research is predictive validity.[18]

Several statistical methods are used for evaluating the predictive validity of risk assessment instruments, such as correlation analysis, use of contingency tables, and receiver operating characteristic (ROC) analysis.[26] Both correlation analysis and contingency tables are affected by the base rate (the frequency that violence occurs in a sample) and thus perform poorly with very low or very high base rates.

ROC analysis is a graphical representation of the trade-off between incorrectly identifying a non-recidivist as a recidivist (false positive) and incorrectly identifying a recidivist as a non-recidivist (false negative). ROC analysis has gained state-of-the-art status in the field for several reasons.[27] The ROC curve may be used to identify suitable cut-off scores for dichotomous decisions. In addition, the area under the ROC curve (AUC) as an estimate of predictive validity is much less sensitive to base rates than are other indices.[27]

An ROC curve is a plot of the true positive rate (sensitivity or 1 - the false negative rate) against the false positive rate (1 - specificity). The area under the ROC curve can be taken as an index for interpreting the overall accuracy of the predictor.[28] Because a perfect risk assessment instrument would have a sensitivity and specificity of one, the closer the area under the ROC curve is to unity, the more accurate the instrument is.[26] An AUC of 0.50 means that the probability that any randomly selected true recidivist would score higher than a randomly selected non-recidivist is 50-50. Thus, an AUC effect size represents a statistically significant improvement over chance if the lower bound of the confidence interval is greater than 0.50, but a statistical improvement over chance alone is not a proper benchmark for practical usefulness. An AUC from 0.60 to 0.69 is usually considered a marginal improvement over chance; an AUC in the range from 0.70 to 0.79 is considered a moderate effect size, and an AUC of 0.80 or above is considered a large effect size.[27] Actual AUC scores obtained in risk assessment research are often above 0.75. In practice, this means that clinicians would accurately (correctly) predict violence in 75% of cases, which represents a sizable improvement over chance.[21]

EVIDENCE

No systematic review that synthesized studies on the predictive validity of risk assessment instruments for spousal violence was found through our comprehensive literature search (see Appendix A: Methodology). A number of clinical reviews discussed the use of various risk assessment instruments for predicting recidivism of spousal violence; however, none of these reviews met the criteria for a systematic review as defined by Cook and colleagues.[29]

Thirty-four primary studies were identified that potentially met our inclusion criteria. On closer examination of the full text articles, eight studies[7,12,18,27,30-33] met the inclusion criteria (see Appendix A: Methodology). Seven studies were published in peer-reviewed journals, whereas one study[12] was published in a report form. Twenty-six studies were excluded and the reasons for exclusion are stated in Appendix B.

Of the eight primary studies, four studies[7,18,32,33] were conducted in Canada, three[12,30,31] in the United States, and one[27] in Sweden. Five studies[12,18,27,30,31] compared the predictive validity of one particular risk assessment instrument with at least one other instrument, whereas three studies[7,32,33] did not. Six risk assessment instruments were primarily evaluated in the included studies (Table 4). Details regarding the study objective, setting, study population, and results were extracted from each primary study and are presented in Appendix C.

Table 4: Characteristics of the risk assessment instruments evaluated in the primary studies

Instrument	Characteristics
Actuarial risk assessment	
Ontario Domestic Assault Risk Assessment (ODARA) Canada	Composition: 13 items, scored 0 or 1 Domains: criminal history, children in the relationship, offender substance abuse, and victim barriers to support Inclusion of psychiatric assessment: no Intended use: criminal justice
Violence Risk Appraisal Guide (VRAG) Canada	Composition: 12 items Domains: demographic, childhood history, and psychiatric assessment variable Inclusion of psychiatric assessment: yes Intended use: courts, clinicians, and parole officials
Structured clinical risk assessment	
Danger Assessment (DA) USA	Composition: 15 (original) to 20 (modified) yes/no items, plus Calendar Assessment of frequency and severity of past-year incidents. Scoring -3 to 37, 4 risk categories (variable, increased, severe, and extreme danger) Domains: escalation of violence, availability of a weapon, generality of the batterer's violence, psychological abusiveness, drug or alcohol abuse Inclusion of psychiatric assessment: no Intended use: victim education, awareness, safety planning, and service provision
Spousal Assault Risk Assessment (SARA) Guide Canada	Composition: 20 items (plus any extra risk factors and critical items considered by the assessor), 3-point scale, total score 0 to 40, summary risk rating (low, moderate, and high) Domains: part 1 (10 items): general violence risk factors and part 2 (10 items): spousal violence risk factors Inclusion of psychiatric assessment: yes Intended use: criminal justice
Domestic Violence Screening Instrument (DVSI) USA	Composition: 12-item checklist, scored 0 to 2 or 0 to 3 for each item, total score 0 to 30 Domains: domestic violence and treatment history, current violence (children present, offender employment, recent separation, and current restraining order or community supervision) Inclusion of psychiatric assessment: yes Intended use: criminal justice
Level of Service Inventory – Revised (LSI-R) Canada	Composition: 3 attitude questionnaires (neutralizations for wife assault, hostile masculinity, and attitudes toward helpers) Domains: criminal history, lifestyle instability, antisocial peers, and antisocial attitudes Inclusion of psychiatric assessment: no Intended use: criminal justice

Studies assessing the reliability/validity of risk assessment instruments

Kropp and Hart[7] conducted a retrospective study to evaluate the reliability and validity of the SARA Guide in judging the risk of violence in adult male offenders. This study included 2,681 males from two groups, 1,671 probationers and 1,010 inmates. However, evaluation of predictive validity of the SARA was based on only 102 males from the probationer group who were convicted of an offence related to spousal violence and who were referred by the criminal courts to attend a treatment program provided in Vancouver, Canada.

In this study, the SARA tool was used by correctional, mental health, and research staff. There was considerable variability in their training and experience. Assessors coded the presence of the 20 risk factors and also indicated whether they considered the risk factor to be a critical item (i.e., the items that were considered to be particularly relevant to the decisions concerning the individual's risk). SARA ratings were based on an interview with the offender and a review of all relevant file information, except for the 102 above-mentioned offenders, for whom researchers calculated ratings on the basis of files only.

The authors noted that inter-rater reliability was low for critical items but high for the total score, Part 1 items, and Part 2 items. Inter-rater reliability was higher for Part 2 items (spousal violence risk factors; ICC = 0.87) than for Part 1 items (general violence risk factors; ICC = 0.68). This difference may be caused in part by the discrepancy of ratings by researchers and correctional staff in items related to mental disorder; that is, researchers evaluated these items more systematically than did correctional staff who in many cases, were forced to rely on existing mental health reports. Furthermore, the inter-rater reliability of summary risk ratings was good, even when the risk ratings were dichotomized (high versus low or moderate risk).

The authors found a clear tendency for those rated as high risk to be recidivists according to official records and for those rated as low risk to be non-recidivists. The AUC of the summary risk ratings was 0.70, indicating that the magnitude of the association between the SARA scores and recidivism was moderate.

As pointed out by the authors, most of the analyses in this study were based on data collected as part of the field implementation of the SARA. There was higher variability in the background and training of assessors in the assessment procedure than one would expect in a controlled research setting. This factor is likely to increase the generalizability of the findings. However, the generalizability to forensic psychiatric populations, female offenders, and agencies outside of North America is unclear. As well, use of implementation data likely resulted in a conservative or lower-bound estimate of the reliability and validity of the SARA ratings. Use of the SARA by assessors with more specific expertise and training should result in improved reliability and validity.

The authors pointed out several limitations of this study. First, inter-rater reliability data were available only for a relatively small sub-sample of offenders. A larger sample might have resulted in higher reliability estimates or at least more stable estimates. Second, the authors were not able to examine the test-retest reliability of the SARA ratings. Third, the authors were unable to examine the usefulness of the SARA for making decisions about risk management.

In a retrospective study, **Hilton and colleagues**[32] examined the predictive accuracy of the Violence Risk Appraisal Guide (VRAG) for a sample of men who had previously committed violent offences against their wives. This instrument was originally developed for predicting recidivism of violence in general.

Eighty-eight male spousal violence offenders who had been admitted to a maximum-security psychiatric facility in the province of Ontario were included in the study. More than half (59%) had an index offence of murder or manslaughter. Of the 88 male offenders, 80 (91%) had an opportunity to recidivate, with a mean time at risk of almost seven years.

In this study, all independent variables were scored by assessors blind to the outcome measures. Inter-rater correlations exceed 0.80, and kappas exceeded 0.70, indicating good inter-rater reliability. In terms of predictive validity, the authors found that the AUC for the VRAG in this sample was 0.75, which closely matched the AUCs of 0.73 to 0.76 for the full VRAG construction and validation samples of 799 men that included these 88 wife assaulters. This result suggests that the VRAG performed as well in predicting recidivism among serious wife assaulters as it has for other serious offenders.

There are several limitations to this study. First, the offenders in this sample were patients in the maximum-security psychiatric hospital and might be considered an unusual sample. Second, this study used only static predictors rather than variables that could change in the weeks, days, and hours before a new offence. Third, more than half of the offenders in this sample had an index offence of murder or manslaughter, ruling out the possibility of predicting repeated violence against the same victim.

In a prospective study, **Hanson and Wallace-Capretta**[33] examined whether the same risk factors commonly associated with criminal recidivism among general offenders are also associated with recidivism among male spousal assaulters. This study was not designed to specifically examine the reliability and validity of one particular risk assessment instrument. The Level of Service Inventory–Revised (LSI–R), originally developed to predict general criminal recidivism, was used, and the predictive validity of this instrument for spousal violence recidivism was reported in this study.

Three hundred and twenty male abusers were recruited from five community treatment programs across Canada. The LSI–R, along with other measures, including family history questionnaire, Locke-Wallace Short Marital Adjustment Test, Expectation of Negative Consequences, Motivation to Change Scale,

Consumer Satisfaction, and Balanced Inventory of Desirable Responding, was administered. The LSI-R combines criminal history, lifestyle instability, antisocial peers, and antisocial attitudes into an overall score. The LSI-R is typically administered in an interview format. However, this study used a self-report version, which was adapted to focus on the problems associated with spousal violence.

The authors found that the predictive validity of the LSI-R total score was in the moderate range, with an AUC of 0.73 for spousal violence recidivism. Inter-rater reliability was not tested in this sample. The authors also noted that the adaptation of the LSI-R for spousal violence offenders was only partially successful.

Studies comparing different risk assessment instruments

Hilton and colleagues[18] developed a brief actuarial assessment instrument, the ODARA, and reported its inter-rater reliability and predictive validity for wife assault recidivism. This is the only published article found that evaluated the ODARA.

The authors identified 689 cases of spousal violence from the Ontario Municipal Provincial Police Automated Cooperative system. Of the 689 cases, 589 served as the construction sample and 100 served as the cross-validation sample. The national Canadian Police Information Centre database was also searched for all records pertaining to each identified offender.

The ODARA evaluates the likelihood that a male who has assaulted his female partner will do so again. The instrument also yields information about offenders' relative rank with respect to this risk, and scores are significantly correlated with the time until subsequent wife assault, its frequency, and its severity. The ODARA did not include measures of psychopathy and other psychological variables known to be strong predictors of violent recidivism by severe spousal assaulters and of violent reoffending in general. The instrument can be easily scored by police officers with no statistical training.

This study demonstrated that the inter-rater reliability of the ODARA was high, with an ICC of 0.90 between researchers and 0.95 between police officers (field simulation). The ROC analysis was used to compare the predictive validity of the ODARA with three other instruments, the Danger Assessment (DA), the SARA Guide, and the Domestic Violence Supplementary Report (DVSR). In the construction sample, the DA, the SARA, and the DVSR all yielded statistically significant predictive validity, but all scored significantly lower than did the ODARA. In the cross-validation sample, which is the more appropriate sample to be used to present predictive accuracy, the ODARA still significantly predicted recidivism, but none of the DA, the SARA, or the DVSR did. However, the predictive accuracy of the ODARA was lower (AUC 0.72) in the cross-validation sample than that in the construction sample (AUC 0.77).

In this study, no offenders committed lethal assaults; thus, no conclusions could be drawn regarding the superiority of the ODARA in predicting lethal recidivism over the DA, which was specifically developed for predicting lethal assault.

The authors concluded that the ODARA offers front-line personnel an easy-to-use actuarial tool that shows promise as a guide for interventions to reduce the incidence of repeated wife assault. The ODARA outperformed other tools in both the construction and cross-validation samples. This study suggests that a substantial increase in predictive validity could be gained by policing services by simply replacing current rational risk assessments with a validated actuarial tool.

In a study conducted in Sweden, **Grann and Wedin**[27] applied the SARA, originally developed in Canada, in a Swedish sample to cumulate cross-cultural experience from the use of the SARA. This study tested the concurrent and predictive validity of the SARA risk factors in 88 offenders convicted of spousal assault or spousal homicide.

The original Canadian version of the SARA was used in this study. Complete evaluation with the SARA required psychological assessment of the perpetrator and clinical judgment.

In this study, the SARA was used as an actuarial tool in a strict sense; that is, the rater did not assign a specific item as 'critical,' no extra factors aside from the 20 items in the SARA were added, and the summary risk rating was simply a numerical addition of the item scores. Because the ratings were based on the information available in the files, some of the items were difficult to assess reliably. In cases in which the file information was insufficient to rate an item, that item was omitted and its score was replaced with the expected mean.

To test inter-rater reliability, a random sub-sample of 18 offenders was rated independently by a graduate-level psychology student and a PhD psychologist. Inter-rater reliability was good for the total scores (ICC = 0.85), but better for the Part 1 items (ICC = 0.88) than for the Part 2 items (ICC = 0.74). The authors did not provide any possible explanation for this difference.

ROC analyses were used to examine the predictive validity of the SARA total scores and the Part 1 and Part 2 scores, respectively. The AUCs ranged from 0.49 to 0.65, indicating that these AUCs correspond to only marginal improvement over chance.

It was noted that in some time frames, the concurrent measures of general violence risk fared better in predicting spousal violence recidivism than did the SARA. For recidivism within 1 year, AUC was higher for the historical part of HCR-20 (H-10) (0.68), for Psychopathy Checklist-Revised (PCL-R) scores (0.71), and for VRAG (0.75) than it was for the SARA total score (0.59).

The ROC analyses suggested that the SARA used in this strictly actuarial manner improved only marginally over chance in terms of statistical prediction of spousal violence recidivism in this sample. The authors found that it was the Part 2 of the SARA (i.e., the items specifically addressing spousal violence risk) that provided the most value to inform risk assessment, although the differences in AUCs between Part 1, Part 2, and total score were small.

The sample in this study was a highly selected one, consisting of offenders who were referred for forensic psychiatric evaluation. The generalizability is not clear. This group of offenders had relatively high mean scores of H-10 and VRAG and was characteristic in terms of previous criminality, suggesting that the group at least in part constituted offenders who exhibit a variety of antisocial traits and behaviours. The sample may thus be offenders for whom spousal violence is a part of an antisocial lifestyle. The fact that the general risk assessment tools (the H-10 and VRAG) predicted recidivism with greater accuracy than the SARA might speak in favour of that hypothesis.

This study had several limitations. An important drawback was its retrospective design and the fact that the risk factors were coded from files. As a result, in some cases one or more of the SARA items had to be omitted because of insufficient information. Poor measurement reliability may have diluted the effect size in the analyses of risk ratios and AUCs. In addition, using reconviction as the criterion variable (as opposed to, e.g. rearrests or self-report) underestimates the true base-rate of recidivism in spousal violence.

Goodman[30] conducted a prospective study using the DA to predict short-term repeated spousal violence. The DA used in this study was the original 15-item instrument compiled from a review of research on risk factors for homicide or serious injury in battering relationships. The DA is short and simple and elicits information solely from the victim. Within many of the settings where domestic violence risk assessment occurs, including the criminal justice system, shelters, and emergency rooms, this feature is an advantage because the victim is usually the only available source for information.

In this study, the initial assessment, including the DA, Revised Conflict Tactics Scales (CTS2), and demographic variables, was conducted in 92 women who presented at the District of Columbia Superior Court following the arrest of their intimate partners who had assaulted them. Three months later, the women were contacted by telephone to ask if any repeated spousal violence occurred during the last three months. Only 49 of 92 women were available at the follow up. Those who were available at follow up and those who were lost to follow up did not differ in their baseline DA, CTS2 scores, or any of the demographic variables.

There were no lethal assaults reported in this study. Using unadjusted and adjusted odds ratios, the authors found that the DA is a much stronger

predictor than the CTS2. No information on inter-rater reliability was available. The authors suggested that, at this point, because research on the DA and other risk assessment instruments is still in an early stage, the scale should be used as a tool for gaining information rather than as a formal predictive instrument.

Campbell and colleagues[12] conducted a prospective multi-site study to test the accuracy of four instruments, the DA, the DV-MOSAIC, the Domestic Violence Screening Instruments (DVSI), and the Kingston Screening Instrument for Domestic Violence (K-SID), in predicting risk of future harm or lethality in spousal violence cases. This study was funded by the US Department of Justice and was published online in a report form (a manuscript of this study was submitted to a peer-reviewed journal and is currently under review, personal communication, Dr. Jacquelyn C. Campbell, March 2007).

Two instruments (the DA and the DV- MOSAIC) were specifically developed for predicting lethal or near lethal reassault, whereas the other two (the DVSI and the K-SID) were developed for predicting repeat violence. The DA used in this study was a modified 20-item instrument.

In this study, 1,307 battered women from family courts, shelters, hospitals, and community offices were allocated to have an interview, using either DA or DV-MOSAIC, and either DVSI or K-SID, to provide the information about their partners. Seven hundred and eighty-two (60%) of them were available for the second interview at the follow up of 5 months to more than 1 year later. Recidivism by the offender was measured by the victim's self-report as well as by arrest data provided by state criminal justice agencies. Information regarding offenders (e.g. attending a batterer's treatment program, reassault, or rearrest) and victims (e.g. safety protective actions) was gathered from the victims' self-report. About 91% of the women who had the follow-up interview took protective actions to reduce the risk of subsequent abuse. However, despite the protective actions, 31% of these women experienced physical or sexual reassault during the time between the baseline and follow up interviews.

All four instruments were found to be significantly, but not highly, related to subsequent severity of abuse. After controlling for the protective actions taken by victims, all of these instruments predicted any assault and severe assault significantly better than chance. The ROC analysis indicated that the AUCs for all four instruments are less than 0.70, with the AUC for DA being higher than that for the other three instruments (see Appendix C). Inter-rater reliability was not reported in this study.

The limitations of this study included high attrition rate (40%), inability to implement the DV-MOSAIC threat assessment system in the way it was designed, and incomplete data collection for the modified version of the DA.

The authors concluded that, without further research, no particular approach can be unequivocally recommended for use in assessing risk in spousal violence cases. The authors suggest that practitioners carefully ask for the victim's perception of her risk, continue to assess risk with all means available, be cautious of potential harm of risk assessment to both victims and perpetrators, use lower risk categories on an instrument to identify cases for intervention if the victim's safety is the greatest concern, and use higher risk categories to identify cases for sanctioning or intensive services if offender fairness or scarce system resources are the greatest concerns.

Williams and Houghton[31] conducted a study using a prospective design to determine the reliability and validity of the DVSI. The 12-item DVSI was designed to be a short statistical tool based on a quick criminal history review that could be made available to prosecutors, judges, and probation officers right after an offender is arrested.

The DVSI was administered to 1,465 male spousal violence offenders to determine the concurrent, discriminant, and predictive validity of the DVSI. To compare the predictive validity of the DVSI with the SARA, both instruments were administered to 434 of 1,465 male offenders because requisite data were available for these 434 males. These male offenders were followed up for 18 months for any repeat spousal violence or other violence.

A quasi-inter-rater reliability assessment suggested that the DVSI was administered reliably in the criminal history record checks.

The ROC analysis based on the 1,465 male offenders showed that AUC for the DVSI was 0.61 for spousal violence reoffending and 0.65 for total reoffending, which can only be seen as a marginal improvement over chance. When comparing the two instruments in the 434 males, the AUC was 0.60 for the DVSI and 0.65 for the SARA for spousal violence reoffending, but the difference was not statistically significant. In other words, the predictive validity for both instruments was similar.

ALBERTA CONTEXT

In September 2005, the Alberta Mental Health Board (AMHB) received annualized funding of $3.995 million from Alberta Health and Wellness for the implementation and enhancement of the Provincial Family Violence Treatment Program in five cities within the Province of Alberta, Canada, including Edmonton, Red Deer, Calgary, Medicine Hat, and Lethbridge. The AMHB develops service agreements with the respective Regional Health Authorities for the communities receiving funding.

The Provincial Family Violence Treatment Program is directed at perpetrators of family violence who are mandated by the court to attend treatment. The Program is accountable to the Mental Health and Justice

Partnering Deputies Committee. A Provincial Family Violence Steering Committee was developed in 2006 to facilitate the coordination, continued development, and delivery of the program. Committee membership includes representatives from ministerial partners, Regional Health Authorities, community treatment providers, and the AMHB.

Since 2005, numerous training and development sessions were hosted by the AMHB to facilitate implementation, including the development of provincial standards. The draft standards, developed collaboratively with stakeholders, include a section on risk assessment. Specifically, the section states:

> Each community shall develop specific risk assessment and management protocols and processes that are part of their Coordinated Community Response. Minimally, these protocols should indicate when and where risk assessments occur, who is responsible for risk assessment, and what processes are to be followed when an offender is considered to be high risk.

Stakeholders have received educational sessions on two risk assessment tools: the SARA and the ODARA. To date, it appears that most treatment providers favour the SARA because it does not require full police reports to complete. The standards do not dictate which risk assessment tools should be used; rather, the document states:

> **Assessment of risk** – a structured risk assessment tool (either the ODARA or the SARA) shall be used prior to treatment and during treatment as necessary since risk is not static. The structured risk assessment tool should be used along with professional/clinical judgement. Participating communities have all received training on the ODARA and the SARA. These should be used in the manner outlined in the manuals. If insufficient information regarding the offender or victim is available to complete the tools, this will affect the validity of results. Efforts should be made to either obtain the information or use another tool.

DISCUSSION

Summary of results

Research on the evaluation of predictive validity of risk assessment instruments is in a very early stage. Problems such as retrospective design, small sample size, failure to fully follow instructions of the instrument when comparing different instruments, and lack of data on lethal or near lethal assault led to inconsistent results in terms of superiority of one instrument over another.

No systematic review on the topic was located. Eight primary studies were found that evaluated several instruments, including the ODARA, the SARA, the DA, the DVSI, the VRAG, and the LSI-R for their predictive validity. The

ODARA, the SARA, and the DA were developed specifically for predicting spousal violence. The ODARA and the SARA were developed in Canada and are currently being used in Canada; the SARA, with some modifications, is also used in other countries. The DA and the SARA are structured clinical risk assessment tools, whereas the ODARA is an actuarial risk assessment tool. The DA was developed and is widely used in the United States. Therefore, the comparison of the predictive validity of these three instruments would provide a better understanding of the current advances in the field.

In terms of predictive validity, the ROC analysis demonstrated that the AUCs for all three instruments are less than 0.80 (see Table 5), indicating that no single instrument has a large effect size. The reported AUCs ranged from 0.59 to 0.77, suggesting a marginal to moderate effect size. In other words, the three instruments predict spousal violence recidivism better than chance.

Table 5: Summary of predictive validity of three risk assessment instruments

Instrument	Approach	ROC-AUC	Improvement over chance
ODARA	Actuarial assessment	0.77 (from one study)[18]	Moderate
SARA	Structured clinical assessment (professional judgment)	0.59 to 0.70 (from four studies)[7,18,27,31]	Marginal
DA	Structural clinical assessment	0.59 to 0.67 (from two studies)[12,18]	Marginal

AUC: area under the curve; DA: Danger Assessment; ODARA: Ontario Domestic Assault Risk Assessment; ROC: receiver operating characteristic

In terms of superiority of the ODARA, the SARA, or the DA, only the study by Hilton[18] provided direct comparison of these three instruments among the same participants. As demonstrated in Table C.1 (Appendix C), the AUC is highest for the ODARA (0.77), followed by the SARA (0.67) and the DA (0.59) in the construction sample. The data analysis indicated that the predictive validity of the ODARA was significantly higher than that of the DA or the SARA. In the cross-validation sample, the ODARA still significantly predicted recidivism, whereas the DA and the SARA did not. However, special caution needs to be taken when interpreting the results from this study. First, when obtaining information for the SARA, the recommended interview and clinical judgment were not available (see previous description about the SARA). Second, no lethal assault was reported in this study; thus, no conclusion can be made that the ODARA is better than the DA in predicting lethal assault. Furthermore, the DA was specifically developed for assessing the risk of lethal or near lethal assault, whereas the ODARA and the SARA were developed for predicting reassault. In the study by Campbell and colleagues[12] that evaluated the predictive validity of the DA and three other instruments, no lethal assault

by male partners was reported. Therefore, there is a lack of available evidence on the predictive validity of risk assessment instruments for lethal assault, which is the greatest concern for female victims.

Inter-rater reliability was tested for the SARA, the ODARA, and the VRAG, but not for the DA and other instruments. All studies demonstrated a good inter-rater reliability for the three instruments. Using intra-class correlation coefficient analysis, two studies[7,27] consistently found a good inter-rater reliability for the total scores of the SARA; however, one study[7] found that the inter-rater reliability was better for Part 2 items (spousal violence risk factors, including factors related to mental disorders) than for Part 1 items (general violence risk factors), whereas the other study[27] reported an opposite finding. One study[7] proposed that systematically different ratings by researchers and correctional staff for items related to mental disorders might have resulted in poorer agreement in Part 1 items; in contrast, although the other study[27] did not offer any explanation, ratings by two individuals with psychology backgrounds may have explained the better agreement in Part 1 items. Future studies need to identify the most appropriate and qualified professional(s) to use risk assessment tools.

Methodological considerations

One of the limitations of this review is that no formal methodological quality appraisal was conducted; however, some methodological issues associated with the included studies were identified.

Sample size

Overall, the sample sizes in the majority of the primary studies are relatively small. Some studies involved a large total number of participants; however, only a small portion of the participants were included for evaluating the predictive validity. The small sample size may affect the determination of the reliability and validity of an instrument. Moreover, it is almost impossible to predict lethal assault, a rare event, in such a small sample size. The difficulty in predicting a relatively rare event such as a homicide necessitates a huge sample size to obtain significant results, even if the instrument predicts accurately.[34] It was estimated that one would need a sample size of more than 50,000 to accurately predict lethality (personal communication, Dr. Janice Roehl, May 2007).

Population characteristics and data collection

The characteristics of the population under study varied considerably across the studies. In some studies, male offenders were arrested, on probation, or in a maximum-security psychiatric facility, whereas in other studies male offenders were referred to attend batterer treatment programs (provided in Canada or the United States). One ongoing criticism of much of the domestic violence research on risk assessment is that it has largely relied on incarcerated individuals, instead of the bulk of perpetrators who appear in court or are

referred for treatment programs without entering the correctional system.[35] The generalizability of the findings from these studies to the current target population in Alberta is questionable.

Two studies[12,30] involved only battered women for obtaining information about their abusive partners. The outcome was determined primarily by women's self report. It has been noted that spousal violence is unique compared with other criminal offences. Because of the relationship, children, and financial dependence, the victim can underestimate and minimize the risk.[6] Male abusers, of course, will significantly minimize their responsibility for violence.[6] Therefore, any risk assessment based only on the accuser's self-report should be made with extreme caution, as the result will likely be underestimated.[6] Ideally, multiple information sources should be used to collect data.

Treatment for abusers and self-protection by victims

In two studies,[7,33] specific treatment programs were provided to the male offenders. Information regarding whether treatment programs were offered to the participants was not available in the other studies. Provision of treatment programs during the follow up period will probably reduce recidivism (assuming the treatment program is effective), and thus may affect the predictive power of the risk assessment instrument under investigation. However, it would be unethical in a prospective study to follow up on participants without offering any interventions.

A unique aspect of prediction in spousal violence is that a particular individual victim is the object of concern. Knowing the identity of the potential victim makes it possible, and therefore incumbent on service providers, to consider her safety as paramount.[12] Furthermore, the known potential target will have her own perception of the dangerousness of the perpetrator. This unique situation creates the need for research on spousal violence risk assessment to take into account the victim's actions as a result of risk assessment and any proactive or preventive response on the part of the victim or criminal justice system, as well as the impact of the treatment interventions provided to the male abuser. These actions or interventions will compromise the ability of the research to detect the accuracy in predicting recurrence of violence, especially lethal or severe reassault.[12]

Campbell and colleagues[12] found that the predictive validity increased when the victim's protective actions during the follow up were taken into account in the analysis. Unfortunately, most primary studies did not conduct a separate analysis or report the results of a separate analysis in which these factors were controlled.

Assessment of mental health status

It is extremely difficult to predict future acts of violence in persons with mental and personality disorders.[36] One reason for this difficulty is that mental illness may not be strongly or directly associated with violence.[36] The general consensus is that assessors should begin their task with a thorough consideration of historical or static factors such as age and previous violence and then turn their analyses to clinical and situational factors.[36]

As shown in Table 4, three instruments (the SARA, the DVSI, and the VRAG) have a psychiatric assessment component, whereas the other three instruments (the ODARA, the DA, and the LSI-R) do not. Conducting a psychiatric assessment may provide additional information about the risk and appropriate treatment plan; however, it makes the instrument lengthy in terms of questions and time, and special training is required for the assessor. Currently, available evidence does not indicate whether including the psychiatric assessment component will increase the predictive validity of the instrument.

Information about the mental health status of participants was available in only two studies;[27,32] therefore, it is difficult to estimate the proportion of male spouse assaulters who have mental disorders.

Outcome measures

There is currently no agreement among researchers and clinicians on an operational definition of risk.[37] The lack of consensus creates several problems. First, it makes it difficult to compare risk assessment studies. Second, it makes it difficult to investigate whether different aspects of risk have correspondingly different constellations of risk indicators. Finally, even if the risk factors are similar for the various forms of spousal violence, it could be that the relative importance or weightings of the risk factors vary.[37]

One of the major reasons for poor predictive validity of interpersonal violence is the assumption that violence is dichotomous and unidimensional.[38] Most studies on spousal violence risk and recidivism appear to define risk in terms of the likelihood that some form of violence will take place sometime in the future. In practice, however, decisions about risk likely involve consideration of the imminence, nature (e.g. emotional, physical, or sexual), frequency, and severity of the violence, in addition to the likelihood that it will occur.[37]

Of the eight primary studies, only two[12,18] analyzed the correlation of the risk assessment instruments with different categories of recidivism severity and found that the predictive validity increased for severe reassault.

Further considerations

In addition to considering current available research evidence, other factors within the local context need to be taken into account in the process of selecting an appropriate risk assessment instrument. These factors may include:

- purpose of using an instrument;
- settings, such as women's shelter, hospital emergency department, community treatment program, mental disorder facility, or justice system;
- users of the instruments, such as clinicians, police officers, treatment program providers, or social workers;
- expertise or experience required in mental health status assessment; and
- availability of information, such as access to the police information systems.

An instrument can be used to assess risk of reoffending or risk of lethal violence. Lethality assessments were designed to be used primarily for prevention, with information primarily from victims. In contrast, most of the instruments designed to assess risk of reassault were developed for use in the criminal justice (or military) system for sentencing, probation, bail, and treatment, with the offender as the primary source of information.[12] For lethal risk assessment, ease of administration and brevity are extremely important. In the criminal justice settings, more actuarial evaluations are feasible and important to conduct and many other sources of data can be obtained.[12]

The strengths of the ODARA are in predicting any reassault but not lethal assaults. The weakness of the ODARA is that several questions require information from criminal justice databases (e.g. previous arrest, previous non-compliance, and previous substance abuse arrest). This information may be available to law enforcement agencies; however, it will frequently not be available to victim advocates, treatment providers, or Child and Family Services (personal communication, Mr. Gary Gibbens, June 2007). The ODARA is based on the actuarial approach; it is intended to estimate the likelihood of future violence rather than to provide information about risk management. This aim means that professionals who use the ODARA still need assistance in making final decisions that reflect the totality of circumstances in the case at hand and that guide case management.[20]

The SARA is aimed at predicting both lethal and reassault outcomes; however, it has been evaluated only on the basis of spousal violence reassault outcomes.[12] Although a few questions could be better answered through access to the criminal justice system database, these gaps do not prevent completion of the risk assessment because of the higher overall number of questions and the critical item structure. Overall, the SARA is a more flexible tool and reflects a more 'clinical' portrait of the abusive individual when compared with the ODARA (personal communication, Mr. Gary Gibbens, June 2007).

The DA emphasizes assessing dangerousness related to lethal, near lethal, or potentially lethal outcomes. The DA is a useful tool in working with victims, because it helps victims begin to look at their own risk and what actions they can take to manage their risk. The DA can be useful to a treatment or law enforcement agency when the abuser cannot be interviewed or there are problems in accessing information. The DA does not measure low-risk or medium violence situations as does the ODARA or the SARA, but it is probably one of the best tools for assessing high risk individuals and providing guidance to potential victims (personal communication, Mr. Gary Gibbens, June 2007).

In summary, the DA may be most appropriately applied by clinicians, victim advocates, or social workers in women's shelters, hospitals, and women's treatment programs. The ODARA, on the other hand, may be more appropriately applied by law enforcement personnel. The SARA may not be an optimal instrument for use by law enforcement personnel because it is relatively long and requires specific judgments about mental health status of the assaulters, such as major mental illness and personality disorders. Thus, completion of the SARA places a relatively heavy burden on users in terms of the availability of time, technical expertise, and case history information.[20] However, the SARA provides an immediate opportunity for the development of a comprehensive treatment plan for the assaulter.

Future research

There is no gold standard definition for spousal violence, and variations exist in definitions across studies. Future research needs to move beyond unidimensional assessment to arrive at definitions of spousal violence that capture multiple domains of aggressive behaviour and allow researchers to reliably classify, clearly communicate, and adequately capture the experience.

Independent evaluation of the inter-rater reliability and predictive validity of an instrument by researchers other than the developers of the instrument is required. This will provide unbiased information about the instrument, which is particularly important for the three instruments (the DA, the ODARA, and the SARA) that were specifically developed to predict spousal or intimate violence.

Future research is required to examine whether including the psychiatric assessment component will increase the predictive validity of the instrument.

Future research needs to correlate risk assessment instruments with different categories of recidivism severity and to take into consideration the treatment interventions provided to the abuser and the self-protective action taken by the victim.

The feasibility, utility, and impact of risk assessment instruments need to be investigated in different settings, relying upon different information sources.

CONCLUSION

Male-to-female spousal violence is a significant public health problem because of the associated physical, psychological, and financial costs to victims, their families, communities, and society in general.

To prevent future serious or lethal assaults and to help allocate already scarce resources, risk assessment instruments have been developed in an attempt to identify spousal violence offenders who are at high risk of reoffending. However, little is known about the reliability and validity of these instruments.

Inter-rater reliability was tested for the SARA, the ODARA, and the VRAG, but not for the DA and other instruments. Limited research indicated good inter-rater reliability for the SARA, the ODARA, and the VRAG.

Five studies compared the predictive validity of one instrument with other instruments, whereas three studies reported predictive validity of only one instrument. Reported ROC AUC (a measure of predictive validity) was less than 0.80 (ranged from 0.59 to 0.77) for all instruments under evaluation, suggesting only marginal to moderate improvement over chance in predicting recidivism.

Only one study that was conducted in Canada compared predictive validity of the ODARA, the SARA, and the DA. On the basis of the ROC analysis, the predictive validity of the ODARA appeared to be lower in the cross-validation sample than that in the construction sample. In the construction sample, all three instruments yielded statistically significant predictive validity, but the ODARA predicted recidivism statistically significantly better than did the SARA or the DA. In the cross-validation sample, the ODARA significantly predicted recidivism, whereas the DA and the SARA did not. However, special caution is needed when interpreting these findings. First, instructions for the SARA were not completely followed; that is, no interview was performed, and no clinical judgment was made. Second, no lethal assault occurred during the follow up in this study; therefore, no conclusion could be made that the ODARA was superior to the DA in predicting lethal assault.

Although lethal assault is of the greatest concern, the included studies failed to provide any information on how well these instruments (even the DA) predict lethal assault. Conducting such studies will be very difficult, if not impossible, because a huge sample size would be needed to detect lethal assault, a rare event.

The decision on selecting an appropriate risk assessment instrument needs to take into account factors such as the available research evidence, the local context, the population under assessment, and the purpose for conducting an assessment. The DA may be most appropriately used by clinicians, victim advocates, or social workers in women's shelters, hospitals, and women's treatment programs. The ODARA, on the other hand, may be more appropriately applied by law enforcement personnel. The SARA may not be an optimal instrument for use by law enforcement personnel

but it provides an immediate opportunity for the development of a more comprehensive treatment plan for the assaulter. More comparative studies on these three instruments in the target population are required.

Currently available research evidence about the predictive validity of risk assessment instruments is very limited. It would be inappropriate to base any decision of an individual's risk solely on the scores of the instruments being used. Information from all other sources such as women's perceptions of risk, or records in the justice system, should be gathered to make more accurate predictions.

The AMHB is in a unique situation where both the SARA and the ODARA are being applied to the target population of interest and within the local context. Data gathered during the implementation of the Provincial Family Violence Treatment Program using these risk assessment tools will provide valuable information regarding their predictive validity and the initiation of a classification system of recidivism severity. This implementation and evaluation process will allow a direct comparison of the predictive validity of the SARA, the ODARA, or a combination of both, in the subgroup of male abusers who were referred to the treatment program, to see whether the combination of the two tools would improve predictive validity. Findings from this process would be helpful to determine whether a composite tool that fits the Alberta context could be developed and used province-wide.

APPENDIX A: METHODOLOGY

Search strategy

The literature search was conducted between 13 October and 14 November 2005, and was updated 28 to 29 May 2007 (Table A.1). Major electronic databases, including PubMed, CRD databases (NHS EED, HTA, DARE), EMBASE, Family & Society Studies Worldwide, Sociological Abstracts, Social Services Abstracts, Social Sciences Abstracts, Academic Search Premier, Web of Science, and PsycINFO, and ABI/INFORM, were searched from January 1995 to May 2007. In addition, relevant library collections and websites of health technology assessment-related agency resources were searched. An Internet search engine was also used to locate grey literature. Search was limited to studies that are published in English or German.

Medical Subject Headings (MeSH) terms relevant to this topic are as follows: domestic violence, spouse abuse, battered women, risk assessment, risk factors, recidivism, homicide, reproducibility of results, reliability, validity, accuracy.

The bibliographies of all included articles were manually searched for relevant references that may have been missed in the database searches.

Table A.1: Summary of search strategy†

Database	Platform	Edition or date searched	Search Terms††
Core Databases			
Cochrane Library	Licenced Resource (Wiley Interface)	May 28, 2007	Title, Abstract or Keywords: (domestic violence or spous* abuse or wife abuse or partner abuse or intimate partner violence)
PubMed	http://www.pubmed.org	May 28, 2007	1. "Domestic Violence"[MeSH:NoExp] or domestic violence [Text Word] or spous* abuse or partner abuse or intimate partner violence or wife abuse 2. tool or tools or test or tests or measure* or instrument or instruments or screening or screen or indicator or indicators or scale or scales or reliability or validity or reproducibility of results or accuracy or predictor or predictors or prediction 3. risk assessment or risk factors or recidivism or homicide or murder 4. (#1 and #2 and #3) 5. (#4) Limit to female 6. (#4) Limit to male 7. #4 not (#6 not #5) Note: #5 – #7 is used to limit to those studies where males are identified, are gender neutral, or have not yet been indexed (excludes studies identified as only about females)

Table A.1: Summary of search strategy† (continued)

Database	Platform	Edition or date searched	Search Terms††
Core Databases (continued)			
CRD Databases (DARE, HTA & NHS EED)	http://www.york.ac.uk/inst/crd/crddatabases.htm	May 28, 2007	domestic violence or spous* abuse or partner abuse or battered women or partner violence
PsycINFO	Licensed Resource (OVID Interface)	May 28, 2007	1. battered females/ or family violence/ or partner abuse/ 2. measurement/ or personality measures/ or psychological assessment/ or psychometrics/ or questionnaires/ or rating scales/ or screening tests/ or standardized tests/ or testing methods/ or test construction/ or prediction/ or (tool$ or instrument$ or test or tests or scale).mp. 3. risk assessment/ or risk factors/ or recidivism/ or homicide/ or risk.mp. 4. 1 and 2 and 3
EMBASE	Licensed Resource (OVID Interface)	May 28, 2007	1. battered woman/ or partner violence/ or battering/ or domestic violence.mp. or partner abuse.mp. or (assault adj2 spous$).mp. or (assault adj2 partner).mp. 2. risk assessment/ or risk factor/ or Recidivism/ or homicide/ 3. prediction/ or screening test/ or psychologic test/ or personality test/ or psychologic assessment/ or rating scale/ or reliability/ or scoring system/ or psychometry/ or questionnaire/ or reproducibility/ or validity.mp. or guide.mp. or tool$.mp. or instrument.mp. or accuracy.mp. 4. 1 and 2 and 3
Web of Science	Licensed Resource (ISI Interface)	May 28, 2007	TS=(domestic violence or spouse abuse or partner abuse or intimate partner violence or wife abuse)and TS=(risk assessment* or risk factor* or recidivism or murder or homicide) and TS=(tool or tools or tests or measure* or instrument or instruments or assessment or screening or screen or indicator* or scale or scales or reliability or validity or reproducibility or accuracy or predictor or prediction)

Table A.1: Summary of search strategy† (continued)

Database	Platform	Edition or date searched	Search Terms††
Core Databases (continued)			
Sociological Abstracts	Licensed Resource (CSA Interface)	May 28, 2007	(KW=("family violence" or "partner abuse" or "spouse abuse" or "domestic violence" or "intimate partner violence")) and (KW=(risk or recidivism or homicide)) and (KW=(accuracy or "measures instruments" or prediction or reliability or scores or tests or validity or tool or tools or screen or screening or scale or scales or indicators))
Social Services Abstracts	Licensed Resource (CSA Interface)	May 28, 2007	(KW=("family violence" or "partner abuse" or "spouse abuse" or "domestic violence" or "intimate partner violence")) and (KW=(risk or recidivism or homicide)) and (KW=(accuracy or "measures instruments" or prediction or reliability or scores or tests or validity or tool or tools or screen or screening or scale or scales or indicators))
Academic Search Premier	Licensed Resource (EBSCO Interface)	May 28, 2007	(SU family violence or domestic violence or abusive men or wife abuse or intimate partner violence or spouse abuse or partner abuse) and (risk assessment or risk factors or homicide or recidivism) and (test or tests or scale or scales or validity or reliability or tool or tools or indicators or prediction or scores or instrument or instruments)
ABI Inform	Licensed Resource (Proquest Interface)	May 28, 2007	(domestic violence) and (risk or murder or homicide or recidivism) and (test* or scale* or validity or reliability or tool* or indicator* or predict* or scores or instrument*)
Family and Society Studies Worldwide	Licensed Resource (Ebsco Interface)	May 28, 2007	(risk assessment or risk factor or risk factors or recidivism or homicide) and (test or tests or scale or scales or validity of reliability or tool or tools or indicators or prediction or predictors or scores or instrument or instruments) and (domestic violence or partner violence or spous* abuse or partner abuse or wife abuse)
Social Sciences Abstracts	Licensed Resource (EBSCO Interface)	May 28, 2007	(domestic violence or abusive men or wife abuse or intimate partner violence or spouse abuse or partner abuse) and (risk or homicide or recidivism) and (SU tests and scales or validity or reliability or tool or tools or indicators or prediction or scores or instrument or instruments)

Table A.1: Summary of search strategy† (continued)

Database	Platform	Edition or date searched	Search Terms††
Core Databases (continued)			
Campbell Collaboration (C2-Ripe)	http://www.campbellcollaboration.org/	May 28, 2007	Browsed list of titles
Instrument Databases			
Mental Measurements Yearbook	Licensed Resource (Ebsco interface)	May 28, 2007	domestic violence or spous* abuse or partner abuse or partner violence or wife abuse
HAPI-Health and Psychosocial Instruments	Licensed Resource (OVID interface)	May 28, 2007	(domestic violence or spous$ abuse or partner abuse or partner violence or wife abuse).mp. and (risk or recidivism or murder or homicide).mp.
Library Catalogues and Theses			
NEOS (Central Alberta Library Consortium)	http://www.library.ualberta.ca/catalogue	May 28, 2007	(domestic violence or spous$ abuse or partner abuse or partner violence or battered women) and (tool? or test? or measure? or instrument? or assessment? or screening or screen? or indicator? or factor? or scale? or reliability or validity or reproducibility of results or accuracy or prediction or predictor?) and (risk or recidivism or murder or homicide)
LocatorPlus (NLM Catalogue) Pubmed interface	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nlmcatalog	May 29, 2007	(domestic violence or spous* abuse or partner abuse or battered women or partner violence) and (validity or reliability or predict* or tool* or test* or instrument* or scale* or assess*)
Proquest Dissertations and Theses	Licensed Resource (Proquest Interface)	May 28, 2007	(domestic violence or spouse abuse or partner abuse or intimate partner violence or wife abuse) and (risk or recidivism or murder or homicide) and (test* or scale* or validity or reliability or tool* or indicator* or predict* or scores or instrument*)

Table A.1: Summary of search strategy† (continued)

Database	Platform	Edition or date searched	Search Terms††
Library Catalogues and Theses (continued)			
Theses Canada Portal	http://www.nlc-bnc.ca/thesescanada	May 28, 2007	Any Keyword: "domestic violence" or "spouse abuse" and risk; Any Keyword: "domestic violence" or "spouse abuse" and recidivism; Any Keyword: "domestic violence" or "spouse abuse" and murder; Any Keyword: "domestic violence" or "spouse abuse" and homicide; Any Keyword: "partner abuse" or "wife abuse" and risk; Any Keyword: "partner abuse" or "wife abuse" and recidivism; Any Keyword: "partner abuse" or "wife abuse" and murder; Any Keyword: "partner abuse" or "wife abuse" and homicide; Any Keyword: "partner violence" and risk; Any Keyword: "partner violence" and recidivism; Any Keyword: "partner violence" and murder; Any Keyword: "partner violence" and homicide;
Grey Literature			
NLH National Library for Health	http://www.library.nhs.uk/	May 29, 2007	Domestic violence; spouse abuse; partner abuse; wife abuse; partner violence Browsed list of results for relevant publications
Other HTA			
AETMIS	http://www.aetmis.gouv.qc.ca/site/home.phtml	May 28, 2007	Browsed list of publications (0 result)
CADTH	http://www.cadth.ca/index.php/en/hta/reports-publications/search	May 28, 2007	Violence; violent; abuse; assault; battered (0 result)
ICES	http://www.ices.on.ca	May 28, 2007	Violence; violent; abuse; assault; battered

Table A.1: Summary of search strategy† (continued)

Database	Platform	Edition or date searched	Search Terms††
Other HTA (continued)			
Health Technology Assessment Unit At McGill	http://www.mcgill.ca/tau/publications/	May 28, 2007	Browsed 2002 – 2007 Reports and Work in progress
Medical Advisory Secretariat	http://www.health.gov.on.ca/english/providers/program/mas/tech/techlist_mn.html	May 28, 2007	Browsed 2001 – 2006 Heatlh Technology Reviews and list of Technology currently under review
CCE	http://www.med. monash.edu.au/healthservices/cce/	May 28, 2007	Browsed list of Critical Appraisals and Full Reports and Completed Projects
ECRI	http://www.ecri.org	May 28, 2007	(domestic violence or spous* abuse or partner abuse or partner violence) and risk; (domestic violence or spous* abuse or partner abuse or intimate partner violence) and recidivism; (domestic violence or spous* abuse or partner abuse or partner violence) and homicide; (domestic violence or spous* abuse or partner abuse or partner violence) and murder;
Health Quality Council, Saskatchewan	http://www.hqc.sk.ca/	May 29, 2007	Browsed list of publications
NICE (UK)	http://www.nice.org.uk/page.aspx?o=ourguidance	May 29, 2007	Browsed Technology Appraisals
NHS Health Technology Assessment Programme	http://www.ncchta.org/	May 29, 2007	Violence; violent; abuse; assault; battered
University HealthSystem Consortium**	http://www.uhc.edu	May 29, 2007	Browsed the publications in print: bulletins, case study; policies and procedures; technology assessments; technology report
NZHTA	http://nzhta.chmeds.ac.nz	May 29, 2007	Browsed publications list

Table A.1: Summary of search strategy[†] (continued)

Database	Platform	Edition or date searched	Search Terms[††]
Government Websites			
Alberta Health and Wellness	http://www.health.gov.ab.ca/resources/publications.html	May 29, 2007	Browsed publications list
Health Canada – National Clearinghouse on Family Violence Library Catalogue	http://www.phac-aspc.gc.ca/ncfv-cnivf/familyviolence/lib_e.html	May 29, 2007	NCFV keyword = diagnostic instruments; NCFV keyword = wife abuse or spouse abuse or abusive men or wife batterers or abused women and NCFV keyword = detection or risk assessment or risk factors or predictors;
Health Canada – National Clearinghouse on Family Violence Publications	http://www.phac-aspc.gc.ca/ncfv-cnivf/familyviolence/publications_e.html	May 29, 2007	Browsed list of publications
Search Engine			
Google	http://www.google. ca	May 29, 2007	With all of the words: validity With at least one of the words: partner-violence Spous-abuse partner-abuse wife-abuse domestic-violence Search within results: risk or recidivism or murder or homicide

Note:

[†] Limits: studies: human studies only. These limits are applied in databases where such functions are available.

[††] "*", "?" and "$" are truncation characters that retrieve all possible suffix variations of the root word e.g. surg* retrieves surgery, surgical, surgeon, etc.

Semi-colons separate searches that were entered separately.

Study selection

The studies identified by the search strategy were retrieved, reviewed, and assessed to determine relevance. This screening and assessment process was conducted by one researcher (BG). Study eligibility was determined according to the predetermined inclusion and exclusion criteria.

Inclusion criteria

Studies were included if they met all of the following criteria:

- Study design: systematic review (defined by Cook et al. 1997[29]), or primary cohort study (prospective or retrospective, with or without comparison of different risk assessment instruments)
- Study focus: evaluating inter-rater reliability and predictive validity of risk assessment instrument(s) for spousal violence recidivism or lethality
- Study population: male spousal violence offenders. No limitation on age.
- Outcome measures: reliability (inter-rater reliability), validity (predictive validity) for spousal violence recidivism (including severe and lethal assault)
- Articles published from 1995 to 2007, full text, in English or German languages

Exclusion criteria

Studies were excluded if they met any one of the following criteria:

- Study design: letters/editorials/news or conference abstracts
- Study focus:
  - screening for onset of spousal violence but not for predicting recidivism
  - prevention or treatment program for spousal violence offenders
- Study population: child abusers, elderly abusers, female spousal violence offenders

Data extraction

Data extraction was performed by one researcher (BG) using a standardized data extraction system developed a priori. The second researcher (CH) reviewed the abstraction tables for accuracy and completeness.

Characteristics of the study

Author(s)

Date of publication

Country where the study was conducted

Study objective(s)

Study population

Total number

Age

Ethnic origin

Characteristics
— Setting
— Marital status
— Education
— Employment
— Substance abuse
— Mental health status
— Access to weapon
— History of violence

Instruments

Full name of the instruments

General description of the instruments
— Development of the instruments
— Composition, domains, and scoring system
— Inclusion of psychiatric assessment
— Intended use
— Source of information

Instrument administrator

Time required for instrument administration

Follow up

Treatment program provided during follow up

Recidivism

Definition

Recidivism rate

Severity

Lethality

Outcome measures

Inter-rater reliability (intra-class correlation coefficient)

Predictive accuracy (odds ratio, ROC AUC, sensitivity, specificity)

Methodological quality appraisal

No formal methodological quality assessment was conducted for the included studies. Some major methodological issues are addressed in the Discussion section.

Expert review

External reviewers with clinical expertise in risk assessment for spousal violence and health technology assessment methodology reviewed the draft report and provided feedback. In selecting reviewers, the practice of the Institute of Health Economics is to choose experts who are well recognized and published in the peer-reviewed literature, and who can offer a good perspective with respect to the use of risk assessment instruments to predict spousal violence recidivism.

APPENDIX B: EXCLUDED STUDIES

Table B.1: Excluded primary studies

Studies	Reason for exclusion
Attala et al. 1994[39]	Did not report predictive validity.
Boris et al. 2002[40]	Did not report predictive validity.
Copas et al. 1996[41]	Did not focus on spousal violence.
Dutton et al. 2001[42]	Did not report predictive validity.
Ellis & Stuckless 2006[43]	Not on recidivism of spousal violence.
Girard & Wormith 2004[44]	Not on recidivism of spousal violence, data on domestic violence group were not separately reported.
Grann et al. 2000[45]	Not on recidivism of spousal violence.
Gray et al. 2004[46]	Not on recidivism of spousal violence.
Harris et al. 1993[47]	Not on recidivism of spousal violence.
Harris et al. 2002[48]	Not on recidivism of spousal violence.
Heckert & Gondolf 2004[49]	Primarily evaluated battered women's perception.
Hendricks et al. 2006[50]	Primarily evaluated 2 treatment programs.
Kroner & Loza 2001[51]	Not on recidivism of spousal violence.
Kroner & Mills 2001[52]	Not on recidivism of spousal violence.
Kropp & Hart 2004[20]	Did not report predictive validity.
Loza & Dhaliwal 1997[53]	Not on recidivism of spousal violence.
Loza et al. 2000[54]	Not on recidivism of spousal violence.
Loza et al. 2000[55]	Not on recidivism of spousal violence.
Loza et al. 2001[56]	Not on recidivism of spousal violence.
Loza et al. 2003[57]	Not on recidivism of spousal violence.
Menzies et al. 1985[58]	Not on recidivism of spousal violence.
Menzies et al. 1994[59]	Not on recidivism of spousal violence.
Nicholls et al. 2004[60]	Not on recidivism of spousal violence.
Stadtland et al. 2005[61]	Not on recidivism of spousal violence.
Weisz et al. 2000[62]	Focused on women's perception.
Williams & Grant 2006[63]	Did not focus on spousal violence.

APPENDIX C: EVIDENCE OF PREDICTIVE ACCURACY OF RISK ASSESSMENT INSTRUMENTS

Abbreviations

AUC – area under the curve
CI – confidence interval
CTS2 – Conflict Tactics Scales (Revised)
DA – Danger Assessment
DSM-III – Diagnostic and Statistical Manual of Mental Disorders, 3rd edition
DV – domestic violence
DVSI – Domestic Violence Screening Instrument
DVSR – Domestic Violence Supplementary Report
FPE – forensic psychiatric evaluation
H-10 – historical part of Historical Clinical Risk-20
HCR-20 – Historical Clinical Risk-20
ICC – intra-class correlation coefficient
K-SID – Kingston Screening Instrument for Domestic Violence
LSI-R – Level of Service Inventory – Revised
NA – not available
NPP – negative predictive power
nss – not statistically significant
ODARA – Ontario Domestic Assault Risk Assessment
PCL-R – Psychopathy Checklist-Revised
PPP – positive prediction power
PPV – positive predictive value
ROC curve – receiver operating characteristic curve
SARA – Spousal Assault Risk Assessment
SD – standard deviation
SE – standard error
SEM – standard error of measurement
Sen. – sensitivity
Spec. – specificity
SV – spousal violence
VRAG – Violence Risk Appraisal Guide

Table C.1: Summary of primary studies

Study	Study population
Kropp & Hart, 2000[7] Canada Retrospective design **Objective:** to evaluate the reliability and validity of judgments concerning the risk of violence using the SARA in adult male offenders	**Total number:** n = 2,681 males **Age:** median 32 years **Ethnic origin:** 80% white, 15% Native Indian, the remainder primarily Asian **Characteristics:** **Setting:** Probationers: 1,671 P1 = 1424 serving terms of probation of up to 2 years P2 = 145 referred by criminal courts to an SV treatment group and serving terms of probation of up to 3 years P3 = 102 convicted of offences related to SV and referred by criminal courts to attend a group treatment program. 52 charged with or convicted of SV after treatment (recidivists), and 50 not. Inmates: 1,010 (65% not for SV) I1 = 552 with a documented history of SV I2 = 86 with a documented history of SV and for whom additional assessment information available I3 = 372 men with no documented history of SV **Marital status:** Most married or lived common-law **Education:** NA **Employment:** NA **Substance abuse:** NA **Mental health status:** NA **Access to weapon:** NA **History of SV:** most
Hilton et al. 2001[32] Canada Retrospective design **Objective:** to examine the predictive accuracy of the VRAG for a sample of men who had previously committed violent offences against their wives	**Total number:** n = 88 males **Age**: mean 37 (SD 12) years **Ethnic origin:** NA **Characteristics:** **Setting:** all had been admitted to the only maximum-security psychiatric facility in Ontario and had faced charges ranging from threatening to first-degree murder following acts of violence against their spouse **Marital status:** 2.3% never married **Education:** NA **Employment:** NA **Substance abuse:** mean alcohol abuse score 1.74 (SD 1.75) **Mental health status:** DSM-III personality disorder 23.9%, DSM-III schizophrenia 13.6% **Access to weapon:** NA **History of SV:** NA

	Instrument	Recidivism	Reliability/validity
	Instrument: SARA **Administrator:** correctional, mental health, and research staff In P1: probation officer In P2: treatment staff (doctoral-level psychologist, counsellors, and social workers) In P3: research assistant (doctoral-level grad student) In I1, I2, I3: correctional staff (case manager) (a grad student coded the SARA in sample I2, blind to the original case manager's rating) **Time required:** NA **Follow up:** NA **Treatment provided:** men in P2 and P3; Men in P3 completed at least 12 to 16 sessions	**Definition:** not clear **Recidivism rate:** 51% in P3, not available for other groups **Severity:** NA **Lethality:** NA **Time to reoffence:** NA	**Inter-rater reliability:** (ICC) (n = 86, from sample I2) **Total score:** 0.84 (P < .05) **Part 1:** 0.68 (P < .05) **Part 2:** 0.87 (P <.05) **Critical items:** 0.22 (P < .05) Summary risk rating **Low vs. moderate. vs. high:** 0.63 (P < .05) **Low/moderate vs. high:** 0.57 (P < .05) **Predictive validity:** **ROC AUC (N = 102, from sample P3):** .70 (SE = .06)
	Instrument: VRAG **Administrator:** NA **Time required:** NA **Follow up:** almost 7 years **Treatment provided:** NA	**Definition:** any new criminal charge for an offence against a person or a readmission to a psychiatric facility for violent behaviour that could have resulted in such a criminal charge **Recidivism rate:** 23.8% **Severity:** NA **Lethality:** none **Time to reoffence:** NA	**Inter-rater reliability:** Inter-rater correlation coefficient > .80 Kappa > .70 **Predictive validity:** **ROC AUC:** .75 (SE = .07)

Table C.1: Summary of primary studies (continued)

Study	Study population
Hanson & Wallace-Capretta 2004[33] Canada Prospective design **Objective:** to examine whether the same risk factors commonly associated with criminal recidivism among general offenders are also associated with recidivism among male batterers	**Total number:** n = 320 males **Age:** mean 35.2 (SD: 8.8) years **Ethnic origin:** NA **Characteristics:** **Setting:** from 5 community treatment programs across Canada **Marital status:** 63.6% legally married to the victim of the index offence **Education:** average 12 years (high school graduation) **Employment:** NA **Substance abuse:** NA **Mental health status:** NA **Access to weapon:** NA **History of SV:** 45.3%
Studies comparing different risk assessment instruments	
Hilton et al. 2004[18] Canada Retrospective design **Objective:** to test predictive validity of both information obtained by officers attending occurrences of male-to-female domestic assault and information maintained in criminal records management systems	**Total number:** n = 689 males (construction sample 589, cross-validation sample100) **Age*:** mean 38 years **Ethnic origin*:** NA **Characteristics*:** **Setting:** not clear **Marital status:** 57% not legally married **Education:** NA **Employment:** 20% unemployment **Substance abuse:** mean score 1.31 (0 to 8) **Mental health status:** NA **Access to weapon:** 9% offence involved weapon **History of SV:** 5% violated prior no-contact order, 3% assaulted victim when pregnant * only for 589 construction sample

** 95% of the time the obtained score would be expected to differ from the true score by ±1.96(.48), or less than 1 point.

	Instrument	Recidivism	Reliability/validity
	Instrument: Level of Service Inventory - Revised (LSI-R) **Administrator:** NA **Time required:** NA **Follow up:** mean 58 months (SD 7.7, range 39 to 73 months) **Treatment provided:** community treatment program	**Definition:** charged or convicted for assault offence, threats, criminal harassment, and harassing telephone calls **Recidivism rate:** 17.2% (spousal violence recidivism) & 25.6% (general recidivism) **Severity:** NA **Lethality:** NA **Time to reoffence:** NA	**Inter-rater reliability:** NA **Predictive validity:** **ROC AUC:** For domestic reoffending: .73 (SD .039) For general reoffending: .76 (SD .31) **Correlation coefficient:** **For spousal violence reoffending:** .32 (P < .001) **For general reoffending:** .40 (P < .001)
	Instrument: ODARA **Administrator:** research assistants **Time required:** NA **Comparator:** DVSR, DA, and SARA **Follow up:** 4.79 years after the index offence (SD = 1.08) **Treatment provided:** NA	**Definition:** any subsequent assault against an (ex)wife or (ex)common-law wife known to police, regardless of whether charges were laid. **Recidivism rate:** 30% in the construction sample and 26% in the cross-validation sample **Severity:** the sum of victim injury scores for all subsequent SV, the sum of Cormier-Lang Scale for all subsequent SV charges, and the number of subsequent domestic incidents with acts of severe violence **Lethality:** none **Time to reoffence:** mean 15.1 months (SD = 12.2) after the index offence	**Inter-rater reliability:** ICC = .90 for ODARA (SEM = .48**) **Predictive validity:** **ROC AUC in the construction sample:** ODARA: .77 (SE = .02, 95% CI = ±.04) DVSR: .67 (95% CI = ±.04) DA: .59 (95% CI = ±.05) SARA: .64 (95% CI = ±.05) (Differences between the ODARA and the other 3 instrument were statistically significant) At ODARA score of 4 Sen. = .59, Spec. = .79, PPP = .54, NPP = .82 **ROC AUC in the cross - validation sample:** ODARA: .72 (SE = .06) DVSR: .59 (95% CI = ±.13) DA: .53 (95% CI = ±.14) SARA: .54 (95% CI = ±.14) **ODARA score and violence severity:** ODARA score positively correlated with 3 measures of severe violence; ODARA score significantly related to the total number of subsequent occurrences of recidivism; ODARA score significantly related to time at risk

Table C.1: Summary of primary studies (continued)

Study	Study population
Grann & Wedin 2002[27] Sweden Retrospective design **Objective:** to test the concurrent and predictive validity of the SARA risk factors in a sample of offenders convicted of spousal assault or spousal homicide	**Total number:** n = 88 males **Age:** mean 34.84 years **Ethnic origin:** 27% 1st generation immigrants **Characteristics:** all committed a violent crime for which they were subjected to forensic psychiatric evaluation (FPE) and clinically diagnosed with a personality disorder **Setting:** 27 (31%) were transferred to forensic psychiatric hospitals, 52 (59%) were sentenced to prison, and 9 (10%) were put on probation. **Marital status:** NA **Education:** 22 (25%) attended college, 15 (17%) did not complete elementary school **Employment:** 43 (49%) unemployed, 46 (52%) financially relied on social welfare **Substance abuse:** 45 (51%) **Mental health status:** 24 (27%) were psychopaths according to the PCL-R **Access to weapon:** NA **History of SV:** 21 (24%) convicted of homicide (murder or manslaughter)

	Instrument	Recidivism	Reliability/validity
	Instrument: Spousal Assault Risk Assessment (SARA) Guide **Administrator:** A graduate-level psychology student (rated all 88 males), and a Ph.D psychologist (rated a random sample of 18 males in order to test inter-rater reliability) **Time required:** NA **Comparator:** PCL-R, historical part of HCR-20 (H-10), VRAG **Follow up:** up to 7.8 yr (mean 3.8, median 4.8, SD 2.4) **Treatment provided:** NA	**Definition:** any reconviction of a hands-off or hands-on violent behaviour, such as homicide, assault, sexual crimes, threats, or violation of no-contact orders, involving a victim with whom the subject had an intimate, sexual relationship. **Recidivism rate:** 28% **Severity:** NA **Lethality:** NA **Time to reoffence:** 15% within 6 months, 18% within 1 year, 24% within 2 years, 25% within 5 years	**Inter-rater reliability (n = 18):** **ICC:** For total score: ICC =.85 (P < .01)] For part 1: ICC =.88 (P < .01)] For part 2: ICC =.74 (P < .01)] **Cohen's kappa:** mean .58 (range .30 to 1.00) for the individual items **Predictive validity:** **SARA individual items and odds ratio (95% CI) for recidivism (1 and 2 vs. 0):** 1) Past violation of conditional release or community supervision: 3.13 (1.26 to 8.71) 2) Personality disorder with anger, impulsivity, or behavioural instability (psychopathy): 7.57 (1.64 to 34.96) 3) extreme minimization or denial of spousal assault history: 8.18 (1.02 to 65.58) **ROC AUC (95% CI):** At 1-yr follow-up: *SARA total score:* .59 (.48 to .70) *PCL-R:* .71 (.60 to .80) *H-10:* .68 (.57 to .78) *VRAG:* .75 (.65 to .84) (statistical significance of the difference was not reported) At 2-yr follow up: *SARA total score:* .63 (.51 to .73) At 5-yr follow up: *SARA total score:* .65 (.51 to .77) *At a cut-off score of 19 (5-yr follow up):* Sen. = .82, Spec. = .50

Table C.1: Summary of primary studies (continued)

Study	Study population
Goodman et al. 2000[30] USA Prospective design **Objective:** to test the degree to which the DA can be used to predict the likelihood of short-term repeat abuse by batterers within the criminal justice system and to determine the relative predictive ability to that of the CTS2	**Total number:** n = 92 women **Age:** older than 18 years **Ethnic origin:** 90% African American **Characteristics:** **Setting:** all women presented at the District of Columbia Superior Court following the arrest of an intimate partner who assaulted them **Marital status:** 12% married to the assaulter, 4% separated from the assaulter **Education:** NA **Employment:** 51% had part- or full-time job **Substance abuse:** NA **Mental health status:** NA **Access to weapon:** NA **History of SV:** NA

*DA and CTS2 were entered simultaneously so that the predictive value of each variable was adjusted for the presence of the other

	Instrument	Recidivism	Reliability/validity
	Instrument: Danger Assessment (DA): 15 items **Administrator:** a clinical psychology graduate student **Time required:** 45 minutes **Comparator:** Revised Conflict Tactics Scales (CTS2): physical assault (12 items), sexual coercion (7 items), and injury (6 items) **Follow up:** 3 months (49 women (53%) available for a follow up phone call) **Treatment provided:** NA	**Definition:** repeat assault (threaten or actual assault) (yes or no) **Recidivism rate:** 22% **Severity:** NA **Lethality:** NA **Time to reoffence:** NA	**Inter-rater reliability:** NA **Predictive validity:** **Unadjusted odds ratios:** DA: 4.18 (95% CI 1.65 to 10.60) ($P < .05$) versus CTS2: 2.77 (95% CI 1.31 to 5.87) ($P < .05$) **Adjusted odds ratios*:** DA: 3.18 (95% CI .61 to 3.82) (nss) versus CTS2: 1.52 (95% CI 1.08 to 9.39) ($P < .05$)

Table C.1: Summary of primary studies (continued)

Study	Study population
Campbell et al. 2005[12] USA Prospective design **Objective:** to test the predictive accuracy of 2 instruments (the DVSI and the K-SID) that assess risk of repeat violence in intimate relationships, and 1 instrument (the DA) and 1 threat assessment method (DV-MOSAIC) assessing lethal or near lethal risk in violent intimate partner relationships	**Total number:** 1,307 women (782 available at follow up interview, retention rate 60%) **Age:** NA **Ethnic origin:** African American 28%, Latina/Hispanic 53%, European descent/White 10%, other 9% **Characteristics:** **Setting:** family courts, shelters, hospitals, and community offices **Marital status:** 28% married or common law **Education:** less than high school 34%, high school 32%, college or vocational school 26%, college degree 9% **Employment:** working full time or part time 44% **Substance abuse:** not applicable **Mental health status:** not applicable **Access to weapon:** not applicable **History of SV:** not applicable

	Instrument	Recidivism	Reliability/validity
	Instrument: DA **Administrator:** researcher **Time required:** 35 to 60 minutes **Comparator:** DV-MOSAIC, DVSI, K-SID **Follow up:** 6 months to 1 year **Treatment provided:** not applicable	**Definition:** 8-point severity – none – verbal abuse – psychological abuse – stalking and threats – physical abuse *Low* (e.g. grabbing, slapping, causing bruises) *Medium* (e.g. kick, slam up against a wall) *High* (e.g. strangling, sexual abuse with force, serious injury) *Very high* (use of a weapon, potentially life-threatening injury) **Recidivism rate:** 84% any types of abuse, 31% physical abuse **Severity:** verbal 6.1%, psychological 30.7%, stalking/threats 16.1%, physical 31.1% (severe 6.3% and very severe 11.3%) **Lethality:** no clearly identified lethal assault **Time to reoffence:** NA	**Inter-rater-reliability:** NA **Predictive validity:** **ROC AUC:** For any physical/sexual abuse: *DA:* .635 (P <.001) *DV-MOSAIC:* .513 *DVSI:* .508 *K-SID:* .551 For severe physical or sexual abuse: *DA:* .67 (P < .001) *DV-MOSAIC:* .589 (P < .05) *DVSI:* .597 (P < .01) *K-SID:* .537 Sen., Spec., and PPV for DA Sen: increased danger .92; severe danger .70; extreme danger .48 Spec. increased danger .22; severe danger .49; extreme danger .68 PPV: increased danger .38; severe danger .42; extreme danger .44 The 4 categories of DA (variable, increased, severe, and extreme danger) highly and significantly associated with the level or type of abuse during the follow up. Scores on DV-MOSAIC and DVSI were also significantly associated with level of abuse at follow up. No statistically significant association between the baseline K-SID risk levels and subsequent abuse.

Table C.1: Summary of primary studies (continued)

Study	Study population
William & Houghton 2004[31] USA Prospective design **Objective:** to use a prospective design to determine the reliability and validity of the Domestic Violence Screening Instrument	**Total number:** n = 1,465 males **Age:** mean 32 years **Ethnic origin:** 49.5% Anglo, 5.6% African American, 43.7% Latino, and 1.2% other. **Characteristics:** **Setting:** arrested for SV offences committed against female partners **Marital Status:** NA **Education:** NA **Employment:** 36% unemployed **Substance abuse:** 39% had prior drug/alcohol treatment **Mental health status:** NA **Access to weapon:** NA **History of SV:** 35% had previous SV convictions and/or restraining order violations

	Instrument	Recidivism	Reliability/validity
	Instrument: Domestic Violence Screening Instrument (DVSI) was administered to 1465 male offenders **Administrator:** probation officer **Time required:** NA **Comparator:** SARA was administered in 434 (out of 1465) male offenders **Follow up:** 18 months **Treatment provided:** NA	**Definition:** arrests for violations of SV restraining orders and arrests for SV reoffending **Recidivism rate:** 29% **Severity:** NA **Lethality:** NA **Time to reoffence:** NA	**Inter-rater reliability:** NA **Predictive validity:** **ROC AUC (at 18-month follow up):** For total 1,465 males: *SV reoffending:* DVSI: .61 (r = .18, P = .00) *Total re-offending:* DVSI: .65 (r = .18, P = .00) For 434 males who were administered both DVSI and SARA: *SV reoffending:* DVSI: .60 (P < .000.) SARA: .65 (P < .000.) Weighted SARA: .65 (P < .000.) Differences among DVSI, SARA, and weighted SARA were not statistically significant *Total reoffending:* DVSI: .68 (P < .000.) SARA: .70 (P < .000.) Weighted SARA: .71 (P < .000.)

References

1. Jamieson, Beals, Lalonde and Associates Inc. *A handbook for health and social service professionals responding to abuse during pregnancy*. Health Canada, Ottawa, ON: The National Clearinghouse on Family Violence; 1999. Available at: http://www.phac-aspc.gc.ca/ncfv-cnivf/familyviolence/pdfs/fem-hndk-pregnancy_e.pdf
2. Cherniak D, Grant L, Mason R, Moore B, Pellizzari R. Intimate partner violence consensus statement. *Journal of Obstetrics and Gynaecology Canada* 2005;157:365-88.
3. Kropp P, Whittemore KE, Hart SC, Webster CD, Eaves D. *Manual for the Spousal Assault Risk Assessment Guide*. Vancouver, BC: The British Columbia Institute on Family Violence; 1995, 2nd Edition.
4. Saunders DG. *Woman battering*. In: Ammerman RTE, Hersen ME, editors. Assessment of family violence: A clinical and legal sourcebook, 2nd Edition. Oxford, England: John Wiley & Sons; 1999: p. 243-70.
5. Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Loazno R, eds. *World report on violence and health: a summary*. World Health Organization, Geneva, Switzerland; 2002. Available at: http://www.who.int/violence_injury_prevention/violence/world_report/en/full_en.pdf.
6. Whittemore KE, Kropp PR. Spousal assault risk assessment: A guide for clinicians. *Journal of Forensic Psychology Practice* 2002;2(2):53-64.
7. Kropp PR, Hart SD. The spousal assault risk assessment (SARA) guide: Reliability and validity in adult male offenders. *Law & Human Behavior* 2000;24(1):101-18.
8. Canadian Centre for Justice Statistics. *Family violence in Canada: a statistical profile 2006*. Ogrodnik L, editor. Ottawa ON: Statistics Canada; 2006. Available at: http://www.statcan.ca/english/freepub/85-224-XIE/85-224-XIE2005000.pdf.
9. Alberta Justice Communications. *Domestic violence handbook for police and crown prosecutors in Alberta*. Edmonton AB: Alberta Government; 2005.
10. Brownridge DA. Male partner violence against Aboriginal women in Canada: an empirical analysis. *Journal of Interpersonal Violence* 2003;18(1):65-83.
11. Tjaden P, Thoennes N. *Prevalence, incidence, and consequences of violence against women: findings from the national violence against women survey*. Washington DC, National Institute of Justice and Center for Disease Control, Research in Brief, November 1998. Available at: http://ncjrs.org/pdffiles/172837.pdf.

12. Campbell JC, O'Sullivan C, Roehl J, Webster D. *Intimate partner violence risk assessment validation study, final report.* National Criminal Justice Reference Service, Office of Justice Programs, U.S. Department of Justice, 2005. Available at: http://www.ncjrs.gov/pdffiles1/nij/grants/209731.pdf 2097312005:i-92.

13. Roehl J, Guertin K. Intimate partner violence: The current use of risk assessments in sentencing offenders. *Justice System Journal* 2000;21(2):171-98.

14. Decker MR. Homicide risk factors among pregnant women abused by their partners: Who leaves the perpetrator and who stays? *Violence Against Women* 2004;10(5):498-513.

15. Kaur G, Herbert L. Recognizing and intervening in intimate partner violence. *Cleveland Clinic Journal of Medicine* 2005;72(5):406-22.

16. Rathus JH, Feindler EL (eds). *Assessment of Partner Violence. A Handbook for Researches and Practitioners.* Washington, DC: American Psychological Association; 2003.

17. Laing L. *Risk assessment in domestic violence.* Australian Domestic and Family Violence Clearinghouse, University of New South Wales, Sydney, Australia, 2004.

18. Hilton N, Harris GT. A Brief Actuarial Assessment for the Prediction of Wife Assault Recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment* 2004;16(3):267-75.

19. Dolan M, Doyle M. Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist. *British Journal of Psychiatry* 2000;177:303-11.

20. Kropp PR, Hart SD. *The development of the brief spousal assault form for the evaluation of risk (B-SAFER): a tool for criminal justice professionals.* Department of Justice Research and Statistics Division, editor. Ottawa, ON: Government of Canada; 2004. Available at: http://canada.justice.gc.ca/en/ps/rs/rep/2005/rr05fv-1/fv-1/pdf.

21. McGuire J. Minimising harm in violence risk assessment: practical solutions to ethical problems? *Health, Risk & Society* 2004;6(4):327-45.

22. Dutton DG, Kropp PR. A review of domestic violence risk instruments. *Trauma and Abuse, a Review Journal* 2000;2(171-2):181.

23. Naughton MJ, Scumaker SA, Anderson RT, Czajkowski SM. *Psychological aspects of health-related quality of life measurement: tests and scales.* In: Spilker B, editor. Quality of Life and Pharmacoeconomics in Clinical Trials, Issue 2. Philadelphia, PA: Lippencott-Raven Publishers; 1996:117-31.

24. Portney LG, Watkins MP. *Foundations of Clinical Research. Applications to Practice.* Mehalik C, editor. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2000, 2nd Edition (Chapters 4, 5, 6: pp. 49-110).

25. Last JM (ed.). *A dictionary of epidemiology.* New York, NY: Oxford University Press; 2001, 4th Edition.

26. Kroner DG. Issues in violent risk assessment: lessons learned and future directions. *Journal of Interpersonal Violence* 2005;20(2):231-5.

27. Grann M. Risk factors for recidivism among spousal assault and spousal homicide offenders. *Psychology, Crime & Law* 2002;8(1):5-23.

28. Douglas KS, Guy LS, Weir J. *HCR-20 violence risk assessment scheme: overview and annotated bibliography (current up to October 3, 2006)*. Burnaby, Canada: Department of Psychology, Simon Fraser University; 2006.

29. Cook DJ, Mulrow CD, Haynes RB. Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine* 1997;126(5):376-80.

30. Goodman LA. Predicting repeat abuse among arrested batterers: Use of the Danger Assessment Scale in the criminal justice system. *Journal of Interpersonal Violence* 2000;15(1):63-74.

31. Williams KR, Houghton AB. Assessing the risk of domestic violence reoffending: A validation study. *Law & Human Behavior* 2004;28(4):437-55.

32. Hilton N, Harris GT, Rice M. Predicting violence by serious wife assaulters. *Journal of Interpersonal Violence* 2001;16(5):408-23.

33. Hanson RK, Wallace-Capretta S. Predictors of criminal recidivism among male batterers. *Psychology, Crime & Law* 2004;10(4):413-27.

34. Campbell JC. *Prediction of homicide of and by battered women.* In: Assessing Dangerousness: Violence by Sexual Offenders, Batterers, and Child Abusers. Campbell JC, editor. Sage Publications, Inc., Thousand Oaks, Calif; 1995, pp. 96-113.

35. Huss MT, Covell CN, Langhinrichsen-Rohling J. Clinical implications for the assessment and treatment of antisocial and psychopathic domestic violence perpetrators. *Journal of Aggression, Maltreatment & Trauma* 2006;13(1):59-85.

36. Webster CD, Douglas KS, Eaves D, Hart SD. *HCR-20: assessing risk for violence, Version 2.* Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University, in cooperation with the British Columbia Forensic Psychiatric Services Commission; 1997.

37. Kropp PR. Some questions regarding spousal assault risk assessment. *Violence Against Women* 2004;10(6):676-97.

38. Campbell JC (ed). *Assessing dangerousness: violence by sexual offenders, batterers, and child abusers.* Sage Publications, Inc., Thousand Oaks, California, 1995.

39. Attala JM, Hudson WW, McSweeney M. A partial validation of two short-form partner abuse scales. *Women's Health* 1994;21(2/3):125-30.

40. Boris NW, Heller SS, Sheperd T, Zeanah CH. Partner violence among homeless young adults: measurement issues and associations. *Journal of Adolescent Health* 2002;30(5):355-63.

41. Copas JB, Marshall P, Tading R. *Predicting reoffending for discretionary conditional release*. Home Office Research and Statistics Directorate, London, UK, Home Office, 1996. Available at: http://www.homeoffice.gov.uk/rds/pdPS2/hors150.pdf.

42. Dutton DG, Landolt MA, Starzomski A, Bodnarchuk. Validation of the propensity for abusiveness scale in diverse male populations. *Journal of Family Violence* 2001;15(1):59-73.

43. Ellis D, Stuckless N. Separation, domestic violence, and divorce mediation. *Conflict Resolution Quarterly* 2006;23(4):461-85.

44. Girard L, Stephen. The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice Behaviour* 2004;31(2):150-81.

45. Grann M, Belfrage H, Tengstrom A. Actuarial assessment of risk for violence. Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice Behaviour* 2000;27(1):97-114.

46. Gray NS, Snowden RJ, MacCulloch S, Phillips H, Taylor J, MacCulloch MJ. Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: a comparative study of HCR-20, PCL:SV, and OGRS. *Journal of Consultant and Clinical Psychology* 2004;72(3):523-30.

47. Harris GT, Rice ME, Quinsey VL. Violent recidivism of mentally disordered offenders. The development of a statistical prediction instrument. *Criminal Justice Behaviour* 1993;20(4):315-35.

48. Harris GT, Rice ME, Cormier CA. Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law & Human Behavior* 2002;26(4):377-94.

49. Heckert DA, Gondolf EW. Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence* 2004;19(7):778-800.

50. Hendricks B, Werner T, Shipway L, Turinetti GJ. Recidivism among spousal abusers: predictions and program evaluation. *Journal of Interpersonal Violence* 2006;21(6):703-16.

51. Kroner DG, Loza W. Evidence for the efficacy of self-report in predicting nonviolent and violent criminal recidivism. *Journal of Interpersonal Violence* 2001;16(2):168-77.

52. Kroner DG, Mills JF. The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice Behaviour* 2001;28(4):471-89.

53. Loza W, Dhaliwal GK. Psychometric evaluation of the risk appraisal guide (RAG). A tool for assessing violent recidivism. *Journal of Interpersonal Violence* 1997;12(6):779-93.

54. Loza W, Dhaliwal G, Kroner EG, Loza-Fanouos A. Reliability, construct, and concurrent validities of the self-appraisal questionnaire. A tool for assessing violent and nonviolent recidivism. *Criminal Justice Behaviour* 2000;27(3):356-74.

55. Loza W, Loza-Fanouos A. Predictive validity of the self-appraisal questionnaire (SAQ). A tool for assessing violent and nonviolent release failures. *Journal of Interpersonal Violence* 2000;15(11):1183-91.

56. Loza W, Loza-Fanouos A. The effectiveness of the self-appraisal questionnaire in predicting offenders' postrelease outcome. A comparison study. *Criminal Justice Behaviour* 2001;28(1):105-21.

57. Loza W, Green K. The self-appraisal questionnaire. A self-report measure for predicting recidivism versus clinician-administered measures: a 5-year follow-up study. *Journal of Interpersonal Violence* 2003;18(7):781-97.

58. Menzies RJ, Webster CD, Sepejak DS. The dimensions of dangerousness. Evaluating the accuracy of psychometric predictions of violence among forensic patients. *Law & Human Behaviour* 1985;9(1):49-70.

59. Menzies R, Webster CD, McMain S, Staley S, Scaglione R. The dimensions of dangerousness revisited. Assessing forensic predictions about violence. *Law & Human Behaviour* 1994;18(1):1-28.

60. Nicholls TL, Ogloff JRP, Douglas KS. Assessing risk for violence among male and female civil psychiatric patients: the HCR-20, PCL:SV, and VSC. *Behavioral Sciences and the Law* 2004;22:127-58.

61. Stadtland C, Hollweg M, Kleindienst N, Dietl J, Reich U, Nedopil N. Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology* 2005;16(1):92-108.

62. Weisz AN, Tolman RM, Saunders DG. Assessing the risk of severe domestic violence: the importance of survivors' predictions. *Journal of Interpersonal Violence* 2000;15(1):75-90.

63. Williams KR, Grant SR. Empirically examining the risk of intimate partner violence: the revised Domestic Violence Screening Instrument (DVSI-R). *Public Health Report* 2006;121(4):400-8.

IHE Publications

- Cost-effectiveness in the detection of syphilis
- The use and benefit of teleoncology services
- Screening newborns for hearing
- Screening newborns for cystic fibrosis
- The use of nitric oxide in acute respiratory distress syndrome
- Routine preoperative tests – are they necessary?
- Consensus statement on self-monitoring in diabetes
- Consensus statement on how to prevent low birth weight
- Evidence of benefit from telemental health: a systematic review
- Economics of childhood immunizations in Canada
- Health technology assessment on the net
- Mental health economic statistics
- World in your pocket
- The use of videoconferencing for mental health services
- Risk assessment tools for predicting spousal violence

This report assesses the research evidence on various risk assessment tools for predicting male-to-female repeat spousal violence and lethality in those males who had contact with the police system.

IHE
INSTITUTE OF
HEALTH ECONOMICS
ALBERTA CANADA

Institute of Health Economics
1200 - 10405 Jasper Avenue
Edmonton AB Canada T5J 3N4

Tel. 780.448.4881 Fax. 780.448.0018
info@ihe.ca

www.ihe.ca

ISBN 978-1-897443-15-6 (print)
ISBN 978-1-897443-16-3 (online)