

**The Validity of
Risk Assessments for
Intimate Partner violence:
A Meta-Analysis**
2007-07

R. Karl Hanson

Leslie Helmus

Guy Bourgon

Public Safety Canada

© Her Majesty the Queen in Right of Canada, 2007

Cat. No.: PS3-1/2007-7

ISBN No.: 978-0-662-05377-4

Authors Note

We would like to thank J. Campbell, L. Cattaneo, Z. Hilton, and R. Kropp for helping us locate studies for this meta-analysis.

The views expressed are those of the authors and are not necessarily those of Public Safety Canada.

Correspondence should be addressed to R. Karl Hanson, Corrections Research, Public Safety Canada, 340 Laurier Ave., West, Ottawa, ON, K1A 0P8

Email: Karl.Hanson@ps-sp.gc.ca

Abstract

This meta-analysis reviews the predictive accuracy of different approaches and tools that are used to assess the risk of recidivism for male spousal assault offenders. In total, 18 studies were found that examined the relationship between an initial assessment of risk and subsequent spousal assault or general violent recidivism. The various approaches to predicting spousal assault recidivism showed, on average, moderate predictive accuracy. The structured tools specifically designed to assess spousal assault risk showed similar levels of accuracy (average weighted d of .40, 10 studies) as tools designed to predict general or violent recidivism (average weighted d of .54, 4 studies) and global assessments of risk provided by the female partners (average weighted d of .36, 5 studies). The most accurate tools were those in which the items were selected empirically (i.e., based on observed predictors in group data). Further research is needed to determine the extent to which the spousal specific risk tools provide useful information not included in the already well-established risk tools designed for general recidivism or violence. Furthermore, it is possible that increased structure could improve the accuracy of the partners' assessment of risk.

Table of Contents

Abstract	i
Introduction.....	1
Method	3
Sample.....	3
Coding procedures	4
Index of predictive accuracy	4
Aggregation of findings	4
Results.....	7
Discussion	9
Implications for practice	11
References.....	13
Table 1 - The average weighted predictive accuracy (<i>d</i>) of various forms of risk assessment for spousal assault offenders.....	18
Table 2 - The average weighted accuracy (<i>d</i>) of individual risk measures for the prediction of spousal assault recidivism	20

Introduction

For a man who has assaulted his intimate partner, a risk assessment can have considerable influence on the responses of police, courts, probation officers and treatment providers. Perhaps most importantly, a risk assessment influences the response and decisions of his victim. Compared to the substantial advances in risk assessment for violent and general criminal recidivism (Andrews & Bonta, 2006, chapter 9; Hanson, 2005), there has been relatively little empirical work on risk assessment for intimate partner violence. Dutton and Kropp's (2000) review described the "science and practice of spousal assault risk assessment [as] still in its infancy" (p. 178). They were hopeful, however, that in the near future, research would find support for some of the various risk assessment tools and procedures that had been proposed. The purpose of the current review is to examine the extent to which that promise has been fulfilled.

The most common approaches to spousal assault risk assessment are a) partner (victim) ratings, b) spousal assault risk scales (both actuarial tools and structured professional judgment), and c) risk scales designed for general or violent recidivism.

In partner ratings, the approach is unstructured judgment. Neither the risk factors nor the method of combining the risk factors into an overall evaluation of risk are specified. It should be noted that this form of assessment is not a singular approach. It is conducted differently by different partners.

In one form of the spousal assault risk scales, evaluators mechanically combine the ratings on a structured list of risk factors into a total score (e.g., DVSI, Williams & Houghton, 2004). Such assessments are often referred to as "actuarial", following Meehl's (1954) classic distinction between actuarial and clinical assessment. In structured professional judgment, evaluators similarly rate a structured list of risk factors, but the overall evaluation of risk is left to professional judgment (e.g., Spousal Assault Risk Assessment [SARA]; Kropp, Hart, Webster, & Eaves, 1995).

Of the spousal assault risk scales, the Dangerous Assessment (DA) scale is the oldest of the measures still commonly used (Campbell, 2005). It was initially developed in the context of emergency room nursing to assess the risk that battered women would be subsequently killed by their partners (Campbell, 1986). Although designed to predict murder (a rare event), it has frequently been used to predict spousal assault recidivism (e.g., Heckert & Gondolf, 2004). Completing the DA is meant to be a collaborative process between the evaluator and the woman who has been a victim of spousal violence. The most recent (2003) version of the scale includes a time line describing the frequency and severity of abuse, 20 yes/no questions (e.g., Does he own a gun? Is he unemployed?) and an algorithm to translate the responses into risk categories (Johns Hopkins University School of Nursing, 2005).

A notable tool is the Spousal Assault Risk Assessment (SARA). The SARA is the most widely used structured judgment tool for spousal risk evaluations. It contains twenty items covering criminal history, psychological functioning, and current social adjustment. The authors emphasize that it is not a test per se, but a guide for structuring professional judgment (Kropp et al., 1995). The quality of the professional judgment is, of course, dependent on the skills and training of the evaluator, as well as the quality of the information available. In research studies, the professional judgment is often bypassed with the final risk rating being based on a simple sum of the risk items (e.g., Williams & Houghton, 2004). When used in this manner, the SARA would be considered a spousal assault risk scale, although it should be noted that this was not the intent of the authors.

Another notable measure is the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004; Mental Health Centre Penetanguishene, 2005), which is classified as a spousal assault risk scale. Unlike many of the other scales in which the items were selected based on theory or prior research, the ODARA was developed empirically. Items that could be reliably assessed by police were examined for their incremental validity in predicting subsequent police contact for spousal assault; next, the scale was tested in a new validation sample. The ODARA contains 13 items, each scored dichotomously and summed to obtain a total score. The items cover substance abuse, the offender's previous history of violence, the number of children in the family, and the victim's barriers to support.

Most of the research on these risk scales is recent. Despite the claims of those who promote particular scales, the most accurate approach to risk assessment has yet to be established. Are these scales more accurate than directly asking the female victims if she expects her partner to assault her in the future? The answer is not yet known. For the prediction of general violence, unstructured opinions about risk have been less accurate than actuarial scales – often barely above chance levels (Quinsey, Harris, Rice & Cormier, 2006, Chapter 4). This trend has also been found with the prediction of general crime recidivism (Andrews & Bonta, 2006). Women's predictions about their partner's violent behaviour may be a special case, however, given their intimate knowledge of the problem (e.g., Weisz, Tolman, & Saunders, 2000).

Another important question is the extent to which special spousal assault risk scales are even necessary. The major risk factors for spousal assault recidivism are similar to the risk factors for general criminal recidivism (e.g., substance abuse, unemployment; Cattaneo & Goodman, 2005; Gendreau, Little, & Goggin, 1996; Hilton & Harris, 2005, in press). As well, several studies have found that the risk scales designed for general and violent recidivism also predicted spousal assault recidivism (Bourgon & Bonta, 2004; Grann & Wedin, 2002; Hendricks, Werner, Shipway, & Turinetti, 2006; Hilton, Harris, Rice, Houghton, & Eke, in press).

As noted previously, the purpose of the present review is to gauge the progress in spousal risk assessment since the Dutton and Kropp (2000) review. Specifically we examine the empirical evidence for the predictive validity of various approaches to risk assessment for male intimate assault offenders. The risk scales were divided into those that were specifically designed to predict spousal assault recidivism (e.g., ODARA, DA), and those designed to predict general or violent recidivism (e.g., Level of Service/Case Management Inventory – Andrews, Bonta, & Wormith, 2004). Although there have been several recent narrative reviews of the research on spousal assault risk assessments (e.g., Campbell, Glass, Sharps, Laughon, & Bloom, 2007; Hilton & Harris, 2005, in press), this would be the first quantitative review using standard meta-analytic techniques (Cooper & Hedges, 1994; Hanson & Broom, 2005).

Method

Sample

Computer searches of PsycINFO, the National Criminal Justice Reference Service (USA), Proquest Digital Dissertations, and Web of Science were conducted using the following key terms: risk assessment, risk instrument, risk scale, prediction, spousal, partner, domestic, wife, marital, assault, abuse, violence, batterers, SARA, ODARA, K-SID, DAS, VRAG, PCL-R, DVSR, PRA, SRA-PA, LSI, PAPS, DV-MOSAIC, PAS. Additional sources included the reference lists of empirical studies and previous reviews, and letters sent to 18 established researchers in the field of spousal assault recidivism.

Studies were included if they examined the ability of risk assessments to predict spousal violence, or any violence (including spousal) recidivism among male offenders released following an index offence for spousal violence. Risk assessments were defined as global assessments of the risk for recidivism (e.g., dangerousness, likelihood of recidivism) made with or without the aid of guidelines or actuarial tools. Studies that only examined specific attributes related to risk (e.g., level of violence, benefit from treatment) were not included. One exception was the PCL-R, which was included because it is occasionally used as a global assessment of risk. The study in this area with by far the largest sample ($N = 14,970$, Williams & Harris, 2006) was excluded because the sample contained a substantial proportion of female offenders (29%).

To be included in this study, risk assessments must have been developed with different samples than those reported in the study (i.e., all tests of risk assessment methods were replications on new samples). All risk assessments were conducted blind to recidivism status. Studies had to include sufficient statistical information to calculate d (the effect size) and the recidivism rate (spousal or violent). For dichotomous variables, at least 5 subjects were needed for all marginal totals.

As of August, 2007, our search yielded 33 usable documents (e.g., published articles, books, government reports, conference presentations). When the same data set was reported in several articles, all the results from these articles were considered to come from the same study. Consequently, the 33 documents represented 18 different studies (country of origin: 10 United States, 6 Canada, and 2 Sweden; 14 (78%) published; produced between 2000 and 2007, with a median of 2003-2004; average sample size of 333, median of 188, range of 49 to 1,465). Most of the offenders were recruited from community settings (10 community, 1 institution, 6 combined, and 1 unknown). When demographic information was presented, the offenders were predominantly Caucasian (8 of 9 studies).

Effect sizes were recorded for two outcome criteria: a) any spousal violence recidivism (versus no recidivism or only non-spousal recidivism - 94 effect sizes); b) any violent recidivism (spousal or non-spousal; versus no recidivism or only non-violent recidivism - 28 effect sizes). Effect sizes for categories of measures were only reported if there were at least three studies.

The most common sources of recidivism information were local (state or provincial) criminal justice records ($k = 9$; 50%) and national records ($k = 9$; 50%). Six studies (33%) used partner report. These percentages add up to more than 100 because some studies used multiple sources. Studies that used some form of criminal justice record ($k = 13$; 72%) used either arrest/charges ($k = 8$), police calls/reports ($k = 3$), or conviction ($k = 2$) as their recidivism criteria. Of the 6 studies that used partner report for

recidivism information, all recidivism criteria had definitions sufficiently narrow that it would qualify as a criminal code offence (e.g., threats to self or extreme jealousy were not considered recidivism). Of the 15 studies (83%) that provided the average follow-up time, the follow-up periods ranged from 2.7 months to 82.5 months, with a mean of 28.5 months ($SD = 24.9$).

Coding procedure

Each study was coded using a standard list of variables and explicit coding rules (available upon request). Eight studies were coded independently by Karl Hanson and Leslie Helmus, and then discussed to develop a consensus. For the first few studies, this process often involved revisions of the coding manual. The remaining 10 studies were coded by Leslie Helmus and the ratings were reviewed by Guy Bourgon. Inter-rater reliability was not formally calculated; however, most coding differences involved simple omissions or clerical errors. Only one finding per individual variable was coded per sample based on sample size, and completeness of information.

Index of predictive accuracy

The effect size indicator was the standardized mean difference, d , defined as follows: $d = (M_1 - M_2)/S_w$, where M_1 is the mean of the deviant group, M_2 is the mean of the non-deviant group, and S_w is the pooled-within standard deviation (Hasselblad & Hedges, 1995). In other words, d measures the average difference between the recidivists and the non-recidivists, and compares this difference to how much recidivists differ from each other, and how much non-recidivists differ from each other.

The d statistic was selected because it is less influenced by recidivism base rates than correlation coefficients – the other statistic commonly used in meta-analyses. According to Cohen (1988), d values of .20 are considered “small”, .50 “medium”, and .80 “large”. The value of d is approximately twice as large as the correlation coefficient calculated from the same data. When the 95% confidence interval for d does not contain zero, it can be considered statistically significant at $p < .05$. When the confidence intervals for two predictors do not overlap, they can be considered significantly different from each other.

Aggregation of findings

Two methods were used to summarize the findings: median values (Slavin, 1995) and weighted mean values (Hedges & Olkin, 1985). The average weighted d value, $d.$, was calculated by weighing each d_i by the inverse of its variance: $d. = \left(\sum_{i=1}^k w_i d_i \right) / \left(\sum_{i=1}^k w_i \right)$, where k is the number of findings, $w_i = 1/v_i$, and v_i is the variance of the individual d_i (fixed effect model). The variance of the weighted mean was used to calculate 95% confidence intervals: $Var(d.) = 1 / \left(\sum_{i=1}^k w_i \right)$; 95% C.I. = $d. \pm 1.96(Var[d.])^{1/2}$. Weighting d values by the inverse of their variance means that findings from small samples are given less weight than findings from large samples.

When d_i was calculated from 2 by 2 tables, the variance of d_i was estimated using Formula 19 from Sánchez-Meca, Marín-Martínez & Chacón-Moscoso (2003), with $1/2$ added to each cell to permit the analysis of tables with empty cells (Fleiss, 1994):

$$\text{Var}(d_i) = .367 \left(\frac{1}{a+.5} + \frac{1}{b+.5} + \frac{1}{c+.5} + \frac{1}{d+.5} \right).$$

When d_i was calculated from other statistics (t , ROC areas, means, etc.), the variance of d_i was estimated using Formula 3 from Hasselblad and Hedges (1995):

$$\text{Var}(d_i) = \left[\frac{N_1 + N_2}{N_1 N_2} + \frac{d_i^2}{2(N_1 + N_2)} \right].$$

To test the generalizability of effects across studies, Hedges and Olkin's (1985) Q statistic was used:

$Q = \sum_{i=1}^k w_i (d_i - d.)^2$. The Q statistic is distributed as a χ^2 with $k-1$ degrees of freedom (k is the number of studies). A significant Q statistic indicates that there is more variability across studies than would be expected by chance. Outliers were excluded from each category if the single extreme value accounted for more than 50% of the total variance (Q).

Results

The observed spousal assault recidivism rate was 28% (1,506/5,338; 14 studies) and the violent recidivism rate (including spousal assault) was 16.4% (280/1,705; 5 studies). The rate of violent recidivism was lower than the rate for spousal assault recidivism because violent recidivism was always based on officially recorded charges and convictions, whereas spousal assault recidivism frequently included more inclusive criteria such as partner reports and police contacts. One study that specified in advance the number of recidivists and non-recidivists was excluded from the rate calculations (Kropp & Hart, 2000). The average follow-up time was 28.5 months. All figures should be considered underestimates because not all offences are reported or sanctioned.

The average weighted predictive accuracy of the various approaches to risk assessment is summarized in Table 1 (see page 18). For the prediction of spousal assault recidivism, the four approaches (spousal assault scales, other risk scales, structured professional judgment, and victim judgment) were similar. The variability within each category was quite low, with the exception of victim judgment, which had one outlying study. Although the differences between the categories were not significant (the confidence intervals overlapped), the risk scales designed to predict other types of recidivism (e.g., criminal, violent) were somewhat more accurate ($d = .54$, 95% C.I. of .42 to .66) than the risk scales designed to predict spousal assault recidivism ($d = .40$, 95% C.I. of .32 to .48). Additionally, structured professional judgment ($d = .36$, 95% C.I. of .19 to .54) and victim judgment ($d = .36$, 95% C.I. of .26 to .45) showed the same accuracy, which was somewhat but not significantly lower than the risk scales designed to predict either spousal assault or other types of recidivism.

For the prediction of violent (including spousal assault) recidivism, only the accuracy of risk scales designed to predict other (e.g., criminal, violent) recidivism is reported because it was the only category with three or more studies. The accuracy of scales designed to predict other types of recidivism was moderate ($d = .63$, 95% C.I. of .48 to .79).

Table 2 presents the weighted predictive accuracy of individual risk measures for the prediction of spousal assault recidivism (see page 20). Measures are organized into two general categories (designed for the prediction of spousal assault versus other types of recidivism), and listed first by the number of validation studies, and then by the sample size in cases where multiple measures have the same number of studies. Within each category, the risk tools showed small to moderate effect sizes, and, with some exceptions, their confidence intervals overlapped. Only one measure had a negative effect size ($d = -.09$; DV-MOSAIC; i.e., offenders deemed to be low risk were actually more likely to re-offend than offenders deemed to be high risk). Interestingly, the two measures with the largest association with spousal assault recidivism were the DVRAG and the VRAG, both of which were developed by the Research Department of the Mental Health Centre in Penetanguishene, Ontario.

To further examine the potential contribution of professional judgment, one of the most commonly used risk scales, the SARA, was divided into studies that formed an overall evaluation of risk based on either a) professional judgment ($k = 2$) or b) summing the items ($k = 5$). The accuracy when SARA items were summed ($d = .43$, 95% C.I. of .32 to .53) appeared somewhat higher than when structured professional judgment was used ($d = .35$, 95% C.I. of .15 to .55), although this difference was not large, the confidence intervals overlapped, and the number of studies was small. It is also worth noting that there was significant variability among the two studies that used SARA to structure professional judgment

($Q = 5.35$, $df = 1$, $p < .05$). Kropp and Hart (2000) found high predictive accuracy ($d = .76$) when the SARA judgements were coded from files by researchers, whereas the predictive accuracy was low ($d = .21$) when the SARA was coded by Swedish police officers in the course of their duties (Kropp, 2003).

Discussion

The present review found moderate predictive accuracy for most of the methods used to predict spousal assault recidivism. Specialized risk scales (actuarial or structured professional judgment) designed for perpetrators of spousal assault showed levels of accuracy similar to the accuracy found for risk scales designed for violent or general recidivism, or the assessment of recidivism risk made by the victims. The lack of evidence concerning the superiority of any one method is likely due to limited research. This meta-analysis was able to identify only 18 studies, all produced since 2000. Consequently, many important research questions could not be addressed. In comparison, there are at least 79 studies of risk assessment for sexual offenders (Hanson & Morton-Bourgon, 2007) and 88 studies published post-1980 examining risk measures and their relationship to violent recidivism (Campbell, French, & Gendreau, 2007).

The equivalence in the average predictive accuracy does not mean that these methods are interchangeable. Different measures could be measuring different constructs based on different information. It is quite possible that there are some risk factors specific to partner assault (e.g., victim's barriers to support) as well as factors relevant to both spousal assault offenders and general offenders (e.g., substance abuse, unemployment). Consequently, it may be possible to improve predictive accuracy by combining specific and general risk factors, as well as by combining information from different sources. The ODARA, for example, includes the victim's assessment of recidivism risk as one of the factors in an actuarial scale completed by police.

Given the limited number of studies, it is too early to identify a specific scale as more accurate than the others in the prediction of intimate partner violence. In comparison to the dozens of replications of individual risk scales for general or sexual recidivism (Andrews, Bonta, & Wormith, 2006; Hanson & Morton-Bourgon, 2007), the spousal violence risk measure that has received the most research ($k = 5$) is a form of the SARA based on adding the items, which is a scoring method contrary to that advocated by the test's developers. Only two studies used the SARA as intended. The next most researched measure ($k = 4$) is Campbell's (2005) DA, which was designed to predict lethality, not spousal assault recidivism. It is interesting to note, however, that the scales showing the strongest relationship with spousal assault recidivism were actuarial measures that were developed empirically (DVRAG, $d = .74$, and VRAG, $d = .65$). The measure with the lowest predictive accuracy (DV-MOSAIC, $d = -.09$) was not designed as a forecasting tool but to assist in "making assessments and case management decisions" (Robert Martin quoted in Berk, He, & Sorenson, 2005).

The history of risk assessment has clearly demonstrated the benefits of structuring risk decisions based on empirical evidence (Quinsey et al., 2006). Not all empirically-based risk factors, however, are equally useful for case management. The most useful risk scales are those that identify the reasons for risk and suggest ways that the risk could be reduced (Andrews et al., 2006). For the assessment of spousal assault risk, there remains considerable opportunity to advance research and practice. More work is needed to identify the specific characteristics of the offenders and their partners that are related to recidivism and amenable to deliberate intervention (i.e., "criminogenic needs").

Another promising approach to spousal assault risk assessment would be to increase the structure of the risk assessments conducted by the partners. To date, the risk assessments of the partners have been responses to single questions (e.g., do you think he will do it again?). Consequently, it is unknown what procedures or information the partners use to determine the risk. Given that increased structure has improved risk prediction in other areas (Andrews et al., 2006; Dawes, Faust, & Meehl, 1989; Hanson & Morton-Bourgon, 2007), it is quite likely that increased structure could also increase the accuracy of the risk assessments conducted by the partners. To date, Campbell's (1986, 2005) DA has been the closest example of this approach; it structures information provided by the partner, but the final risk rating is made by the evaluator, not the partner, who may or may not agree. Although structured risk assessment by partners is an important direction for research, it has limitations, such as requiring the partner's cooperation (Lewin, Strand, & Belfrage, 2007) and the possibility of the victim's actions impacting their assessment (e.g., "there's no risk for re-offence because I am leaving him...").

Implications for practice

Before deciding which risk assessment approach to use, evaluators need to understand the purpose for the assessment. Some assessments focus on the victim's needs for protection and assistance; others focus on the offender's likelihood of re-offending. It is important to note that none of the scales examined in the current review directly address questions of whether the female partner needs help, nor whether the relationship should continue.

This study found the victims' assessment of risk to have similar levels of predictive accuracy to the other approaches to risk assessment. Given that such assessments are credible and cost-effective, they should be considered wherever possible. Further research is needed to determine whether the victims' judgment can be improved through increased structure, and, if so, how to combine the partners' judgment of risk with other risk relevant information.

For standard correctional practice and supervision, perpetrators of spousal assault could be assessed using risk tools designed for general or violent recidivism. The general risk assessment tools perform as well as the specialized spousal assault tools in predicting spousal assault recidivism, and further research is needed before we know if the specialized tools contain risk relevant information not contained in the other scales (i.e., incremental validity). Case managers, however, may still want to consider some of the items in the specialized tools as a guide to interventions (e.g., partner's barriers to support). For the purpose of pre-treatment assessment, the domestic-specific scales could have utility if they identify appropriate criminogenic needs. Evaluators should be cautious in the interpretation of these results because the extent to which any of these scales assess the unique criminogenic needs of spousal assault offenders has yet to be established.

References

Studies with an asterisk were included in the meta-analysis.

- Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Cincinnati, OH: Anderson.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *The level of service/case management inventory (LS/CMI)*. Toronto, Ontario: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52, 7-27.
- Berk, R. A., He, V., & Sorenson, S. B. (2005). Developing a practical forecasting screener for domestic violence incidents. *Evaluation Review*, 29, 358-383.
- *Bourgon, G., & Bonta, J. (2004). *Risk assessment for general assault and partner abusers* (User Report No. 2004-04). Ottawa, ON: Public Safety and Emergency Preparedness Canada.
- Campbell, J. C. (1986). Nursing assessment for risk of homicide with battered women. *Advances in Nursing Science*, 8, 36-51.
- Campbell, J. C. (2005). Assessing dangerousness in domestic violence cases: History, challenges, and opportunities. *Criminology and Public Policy*, 4, 653-672.
- *Campbell, J. C., O'Sullivan, C., Roehl, J., & Webster, D. (2005). *Intimate partner violence risk assessment validation study, final report*. Washington, DC: National Institute of Justice.
- Campbell, J. C., Glass, N., Sharps, P. W., Laughon, K., & Bloom, T. (2007). Intimate partner homicide: Review and implications of research and policy. *Trauma, Violence, & Abuse*, 8, 246-269.
- Campbell, M. A., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism for adult offenders* (User Report No. 2007-04). Ottawa, ON: Public Safety Canada.
- *Cattaneo, L. B., Bell, M. E., Goodman, L. A., & Dutton, M. A. (in press). Intimate partner violence victims' accuracy in assessing their risk of re-abuse. *Journal of Family Violence*.
- *Cattaneo, L. B., & Goodman, L. A. (2003). Victim-reported risk factors for continued abusive behaviour: Assessing the dangerousness of arrested batterers. *Journal of Community Psychology*, 31, 349-369.
- Cattaneo, L. B., & Goodman, L.A. (2005). Risk factors for reabuse in intimate partner violence: A cross-disciplinary critical review. *Trauma, Violence & Abuse*, 6, 141-175.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Cooper, H. & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668-1674.
- *Dutton, D. G., Bodnarchuk, M., Kropp, R., Hart, S. D., & Ogloff, J. R. P. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. *International Journal of Offender Therapy and Comparative Criminology*, *41*, 9-23.
- Dutton, D. G., & Kropp, P. R. (2000). A review of domestic violence risk instruments. *Trauma, Violence & Abuse*, *1*, 171-181.
- Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Russell Sage Foundation.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, *34*, 575-607.
- *Girard, L. (1999). *The Level of Service Inventory-Ontario Revision: Risk/need assessment and recidivism*. Unpublished doctoral dissertation. University of Ottawa.
- *Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior*, *31*, 150-181.
- *Glackman, W. (2004). *Spousal Assault Risk Assessment (SARA) follow-up study: Overall rating and reoffence*. Victoria, BC: Unpublished report, Ministry of Public Safety and the Solicitor General.
- *Gondolf, E. W. (2002). *Batterer intervention systems*. Thousand Oaks, CA: Sage.
- *Goodman, L. A., Dutton, M. A., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers: Use of the Danger Assessment Scale in the criminal justice system. *Journal of Interpersonal Violence*, *15*, 63-74.
- *Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behaviour*, *27*, 97-114.
- *Grann, M., Langstrom, N., Tengstrom, A., & Kullgren, G. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. *Law and Human Behavior*, *23*, 205-217.
- *Grann, M., & Wedin, M. (2002). Risk factors for recidivism among spousal assault and spousal homicide offenders. *Psychology, Crime & Law*, *8*, 5-23.
- Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, *20*, 212-217.

- Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: Application to programs for reducing sexual violence. *Sexual Abuse: A Journal of Research and Treatment, 17*, 357-373.
- Hanson, R. K., & Morton-Bourgon (2007). *The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis* (User Report No. 2007-01). Ottawa, ON: Public Safety Canada.
- *Hanson, R. K., & Wallace-Capretta, S. (2004). Predictors of criminal recidivism among male batterers. *Psychology, Crime, & Law, 10*(4), 413-427.
- *Hanson, R. K., & Wallace-Capretta, S. (2000). *A multi-site study of treatment for abusive men* (User Report No. 2000-05). Ottawa, ON: Solicitor General Canada.
- *Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior, 15*, 625-637.
- Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117*, 167-178.
- *Heckert, D. A., & Gondolf, E. W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence, 19*, 778-800.
- *Heckert, D. A., & Gondolf, E. W. (2005). Do multiple outcomes and conditional factors improve prediction of batterer reassault? *Violence and Victims, 20*, 3-24.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- *Hendricks, B., Werner, T., Shipway, L., & Turinetti, G. J. (2006). Recidivism among spousal abusers: Predictions and program evaluation. *Journal of Interpersonal Violence, 21*, 703-716.
- *Hilton, N. Z. (2006). [Greater Toronto Area police databases]. Unpublished data (email correspondence).
- Hilton, N. Z., & Harris, G. T. (2005). Predicting wife assault: A critical review and implications for policy and practice. *Trauma, Violence & Abuse, 6*, 3-23.
- Hilton, N. Z. & Harris, G. T. (in press). Assessing risk of intimate partner violence. In J. C. Campbell (Ed.), *Assessing dangerousness: Violence by sexual offender, batterers and child abusers* (2nd ed.). New York: Springer.
- *Hilton, N. Z., Harris, G. T., & Rice, M. E. (2001). Predicting violence by serious wife assaulters. *Journal of Interpersonal Violence, 16*, 408-423.
- *Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (in press). An indepth actuarial assessment for wife assault recidivism: The *Domestic Violence Risk Appraisal Guide*. *Law & Human Behavior*.
- *Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment, 16*, 267-275.

- Johns Hopkins University School of Nursing. (2005). *Danger Assessment*. Downloaded September 6, 2007 from <http://www.dangerassessment.org/WebApplication1/pages/product.aspx>.
- *Kropp, P. R. (2003, April). *Validity of law enforcement officers evaluations of risk*. Paper presented at the 3rd Annual Conference of the International Association of Forensic Mental Health Services, Miami, Florida.
- *Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior, 24*, 101-118.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide* (2nd ed.). Vancouver, BC: British Columbia Institute on Family Violence.
- Lewin, C., Strand, S., & Belfrage, H. (2007, June). *Victims not cooperating in police investigations: The case of spousal assault. How can they be persuaded to cooperate?* Paper presented at the 7th annual conference of the International Association for Forensic Mental Health Services, Montreal, Ontario, Canada.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Mental Health Centre Penetanguishene. (2005). *Ontario Domestic Assault Risk Assessment: General Scoring Criteria*. Penetanguishene, ON: Author.
- *Murphy, C. M., Morrel, T. M., Elliott, J. D., & Neavins, T. M. (2003). A prognostic indicator scale for the treatment of partner abuse perpetrators. *Journal of Interpersonal Violence, 18*, 1087-1105.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd Ed.). Washington, DC: American Psychological Association.
- *Shepard, M. F., Falk, D. R., & Elliott, B. A. (2002). Enhancing coordinated community responses to reduce recidivism in cases of domestic violence. *Journal of Interpersonal Violence, 17*, 551-569.
- Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*, 448-467.
- Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*, 9-18.
- *Weisz, A. N., Tolman, R. M., & Saunders, D. G. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence, 15*, 75-90.
- Williams, K. R., & Grant, S. R. (2006). Empirically examining the risk of intimate partner violence: The Revised Domestic Violence Screening Instrument (DVSI-R). *Public Health Reports, 121*, 400-408.
- *Williams, K. R., & Houghton, A. B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior, 28*, 437-455.

Table 1. The average weighted predictive accuracy (*d*) of various forms of risk assessment for spousal assault offenders

Variable	Median	Mean	95% C. I.		<u>Q</u>	<i>k</i>	<i>n</i>	Studies
Spousal Assault Recidivism								
Spousal Assault Scales	.45	.40	.32	.48	13.86	10	3,268	Bourgon & Bonta (2004); Campbell et al. (2005); Goodman et al (2000); Grann & Wedin (2002); Heckert & Gondolf (2004); Hilton et al., 2004; Hilton et al. (in press); Kropp & Hart (2000); Murphy et al. (2003); Williams & Houghton (2004).
Other Risk Scales	.52	.54	.42	.66	4.16	4	1,438	Bourgon & Bonta (2004); Grann & Wedin (2002); Hendriks et al (2006); Hilton et al. (in press).
Structured Professional Judgment	.40	.36	.19	.54	5.41	3	658	Kropp (2003); Kropp & Hart (2000); Shepard et al. (2002).
Victim Judgment	.47	.36	.26	.45	5.65	5	2,179	Campbell et al. (2005); Cattaneo et al. (2006); Cattaneo & Goodman (2003); Heckert & Gondolf (2004); Hilton et al. (2004).
With Weisz et al. (2000)	.49	.40	.31	.49	21.19***	6	2,356	

Table 1 continued

Variable	Median	Mean	95% C. I.	<i>Q</i>	<i>k</i>	<i>n</i>	Studies
Any Violent Recidivism							
Other Risk Scales	.74	.63	.48 .79	6.75	4	1,039	Bourgon & Bonta (2004); Girard & Wormith (2004); Hanson & Wallace-Capretta (2004); Hilton et al. (2001).

Table 2. The average weighted accuracy (*d*) of individual risk measures for the prediction of spousal assault recidivism

Variable	Median	Mean	95% C. I.		<u>Q</u>	<i>k</i>	<i>n</i>	Studies
Designed for Spousal Assault								
SARA – Total Score	.47	.43	.32	.53	3.60	5	1,768	Grann & Wedin (2002); Heckert & Gondolf (2004); Hilton et al. (2004); Kropp & Hart (2000); Williams & Houghton (2004).
DA	.58	.41	.31	.52	18.47***	4	1,585	Campbell et al. (2005); Goodman et al. (2000); Heckert & Gondolf (2004); Hilton et al. (in press).
DVSI	.39	.33	.24	.41	12.71**	3	2,487	Campbell et al. (2005); Hilton et al. (in press); Williams & Houghton (2004).
KSID	.14	.15	.00	.30	1.95	2	881	Campbell et al. (2005); Heckert & Gondolf (2004).
DVSR	.47	.58	.41	.75	1.49	2	689	Hilton et al. (2004); Hilton et al. (in press).
SARA – Structured Professional Judgment	.48	.35	.15	.55	5.35*	2	531	Kropp (2003); Kropp & Hart (2000).
ODARA	.68	.60	.40	.79	1.13	2	446	Hilton et al. (2004); Hilton et al. (in press).

Table 2 continued

Variable	Median	Mean	95% C. I.		<u>Q</u>	<i>k</i>	<i>n</i>	Studies
SRA-PA		.39	.13	.65		1	502	Bourgon & Bonta (2004).
DVMOSAIC		-.09	-.31	.12		1	367	Campbell et al. (2005).
DVRAG		.74	.52	.96		1	346	Hilton et al. (in press).
EDAIP		.40	.05	.76		1	127	Shepard et al. (2002).
PAPS		.62	-.02	1.25		1	67	Murphy et al. (2003).
Designed for Other Recidivism								
VRAG	.78	.65	.49	.80	1.23	2	736	Grann & Wedin (2002); Hilton et al. (in press).
PCL-R	.68	.60	.45	.75	.46	2	736	Grann & Wedin (2002); Hilton et al. (in press).
PRA		.36	.10	.62		1	502	Bourgon & Bonta (2004).
LSI-R		.43	.06	.79		1	200	Hendricks et al. (2006).