A Brief Actuarial Assessment for the Prediction of Wife Assault Recidivism: The ODARA

N. Zoe Hilton, Grant T. Harris, Marnie E. Rice. Carol Lang, Catherine A. Cormier
Penetanguishene Mental Health Centre
and
Kathryn J. Lines
Ontario Provincial Police

Running Head: WIFE ASSAULT RISK ASSESSMENT

A Brief Actuarial Assessment for the Prediction of Wife Assault Recidivism: The ODARA

N. Zoe Hilton, Grant T. Harris, Marnie E. Rice. Carol Lang, Catherine A. Cormier

Penetanguishene Mental Health Centre

and

Kathryn J. Lines

Ontario Provincial Police

Abstract

An actuarial assessment to predict male-to-female marital violence was constructed from a pool of potential predictors in a sample of 589 offenders identified in police records, followed up for almost five years on average. Archival information in several domains (offender characteristics, domestic violence history, nondomestic criminal history, relationship characteristics, victim characteristics, index offense) and recidivism was subjected to set-wise and step-wise logistic regression. The resulting 13-item scale (Ontario Domestic Assault Risk Assessment, ODARA) showed a large effect size in predicting new assaults against legal or common-law wives or ex-wives, Cohen's d = 1.1, ROC area = .77, and was associated with number and severity of new assaults and time until recidivism. Cross-validation and comparison with other instruments are also reported.

A Brief Actuarial Assessment for the Prediction of Wife Assault Recidivism: The ODARA

When a man who is already known to authorities as a wife assaulter kills his partner, the public demands to know why the criminal justice system did not protect this woman. The need for authorities to recognize the danger using standard risk assessments for repeated assault has been identified (e.g., Office of the Chief Coroner of Ontario, 1998; 2002), providing a timely opportunity for psychological assessment research to make an important contribution. A useful risk assessment would accurately appraise the likelihood that a man who has just assaulted his wife would do so again if he has the opportunity. Accused men obtaining the highest score on a valid risk assessment could, for example, be considered suitable for pre-trial detention by police or bail courts, and less suitable for reliance on protection orders

Formal risk assessments for criminal violence have appeared in the psychological literature (e.g., Harris, Rice, Quinsey, Lalumiere, Boer, & Lang, 2003; Rice, 1997; Rice & Harris, 1995; 2002) and have been used successfully to assess and manage violent offenders (e.g., Quinsey, Harris, Rice, & Cormier, 1998). Although this literature uses the techniques of psychological test construction, the resulting risk assessments are not psychological tests *per se*. That is, although creating such an assessment device requires the empirical identification of relevant items, efficient combination of items, development of norms, and demonstration of reliability and validity, the primary test of risk assessment research is predictive validity. Formal risk assessments are not designed to measure an underlying hypothetical psychological construct or disposition. Rather, their principal purpose is to estimate the likelihood of overt behavior, such as interpersonal violence, often measured imperfectly by arrest or conviction for a violent crime.

Consequently, item selection by factor analysis, internal consistency, test-retest reliability, and construct validity, though central to psychological test construction, are less central to formal risk assessment construction. The latter is characterized by using multiple regression to select items, establishing inter-rater reliability, and demonstrating predictive validity and cross validation.

Only in the past 10-15 years have predictor items for wife assault recidivism been empirically identified. Individual and social variables consistently related to wife assault recidivism include age, severity and duration of prior violence, other prior antisocial behavior, violence in the offender's family of origin, hostility, and substance abuse (e.g., Aldarondo & Sugarman, 1996; Dutton, Bodnarchuk, Kropp, Hart, & Ogloff, 1997; Hilton, Harris, & Rice, 2001; Saunders, 1993; Shepard, 1992). This research has not previously been extended to the selection of items for risk assessment using regression techniques. Literature reviews, however, have inspired structured lists for assessing risk among wife assaulters (see reviews by Dutton & Kropp, 2000; Roehl & Guertin, 2000). The *Danger Assessment* (DA; Campbell, 1986; Stuart & Campbell, 1989) is one such structured clinical assessment. It was designed to assess the risk of lethal wife assault using victim interview or self report. Its 15 items pertain to a perpetrator's history of relationship and other violence, availability of weapons, substance abuse, suicidality, and jealousy. The sum of DA items exhibited test-retest reliability (*r*s > .83; Campbell, 1995; inter-rater reliability has not been reported), and was positively associated with past physical or sexual victimization (Campbell, Soeken, McFarlane, & Parker, 1998; McFarlane, Parker, & Soeken, 1995). A small positive association was also reported between DA score and subsequent nonlethal violence or serious threats of violence four months after the disposition of a domestic court (Weisz, Tolman, & Saunders, 2000) and between several DA items and nonlethal assault or

threats three months after the arrest of a perpetrator (Goodman, Dutton, & Bennett, 2000).The predictive value of the DA might be limited by items that are actually inversely associated with recidivism (e.g., suicidality; Hilton et al., 2001), which could be avoided by using empirical item selection.

The *Spousal Assault Risk Assessment* (SARA; Kropp, Hart, Webster, & Eaves, 1995) was informed by the empirical literature and clinical interpretation of variables that distinguish domestically violent men. Part 1 has ten "general violence risk factors" which include substance abuse, employment problems, mental illness, and personality disorder, and items that may reflect domestic conflict including suicidal/homicidal relationship problems, exposure to family violence, violation of conditional release, and past assault of family members. Part 2 has ten "spousal violence risk factors" including recent escalation, offender attitudes, and characteristics of the most recent assault. The manual instructs assessors to use the items, especially those judged to be critical, to form a clinical judgment of low, moderate, or high risk. Interrater reliability was not good for either critical items or the clinical risk rating (ICC = .18 to .63; Kropp & Hart, 2000). An unadjusted total of the scores on the 20 items exhibited better inter-rater reliability (.84) but was unrelated to wife assault recidivism (Kropp & Hart, 2000). Grann and Wedin (2002) did find significant prediction of wife assault recidivism using the SARA total, depending on the follow-up time. The risk rating and scores on Part 2 were higher among wife assault recidivists than nonrecidivists, at some follow-up intervals (Grann & Wedin, 2002; Kropp & Hart, 2000). As with the DA, it appears that only some SARA items are useful predictors; i.e., violation of conditional release, personality disorder (including psychopathy), and minimization or denial of spousal assault (Grann & Wedin, 2002). Weak, inverse, or unreliable

predictors are less likely to be selected by empirical test construction methods.

Psychological researchers have developed actuarial methods to appraise the risk of violence posed by offenders in general. For example, the *Violence Risk Appraisal Guide* (VRAG; Harris, Rice, & Quinsey, 1993; Quinsey et al., 1998) was developed using the test construction methods described above. It has been cross-validated with large predictive effect sizes in many samples of violent male offenders (eg., Glover, Nicholson, Hemmati, Bernfeld, & Quinsey, 2002; Harris & Rice, 2003; Harris, Rice, & Camilleri, in press; Harris, Rice, & Cormier, 2002; Rice & Harris, 1997). Its 12 items include demographic, childhood history, criminal offense, and psychiatric variables. When scored according to its published instructions (Quinsey et al., 1998), the VRAG exhibits high inter-rater reliability ($r > .90$). VRAG scores are positively related to the likelihood and severity of violent reoffending and inversely to time until violent recidivism (Harris et al., 2002; Harris et al., 2003). The VRAG performed equally well in predicting violent recidivism within ten years among 81 men at risk after assaulting their wives as it did for the full original development sample (Hilton et al., 2001), but it was unknown whether the victims of the subsequent were domestic partners. Grann and Wedin (2002) reported a large effect for the VRAG's one-year predictions of wife assault recidivism in 88 personality disordered men. More problematic, however, is that the VRAG requires extensive knowledge of the offender's life history and psychological characteristics, so is not suitable for rapid risk assessment by police officers or courts dealing with domestic violence cases. The present study attempted to meet the need for an assessment that can be quickly completed using only the information readily available to these users.

In our jurisdiction (Ontario, Canada) current practice requires police officers to complete

a documented domestic violence investigation including a victim interview and the perpetrator's criminal record accessed from an automated criminal records system. The investigation is used to complete the *Domestic Violence Supplementary Report* (DVSR). Based on a literature review and consultation with experts, the DVSR was developed by the Ontario Provincial Police in response to highly publicized murder-suicides by men already known to the police for wife assault (Ministry of the Solicitor General, 2000). Its 22 items pertain to relationship separation, jealousy, stalking, threats, weapons, substance abuse, mental illness, disobeying court orders, violence towards pets, and the victim's fear, but not general criminal history. An advantage of the DVSR for law enforcement is that, unlike the DA or SARA, it was specifically designed for front-line officers and relies only on information readily available when investing domestic violence incidents. There are no data about the predictive accuracy of the DVSR.

As mentioned earlier, the development of the DA, SARA, and DVSR did not benefit from the science of psychological assessment in general, or the use of actuarial methods in particular. "Actuarial" here refers to the selection of predictor items based on their measured association with outcomes in representative samples and the combination of predictors based on incremental validity. It has been well established that actuarial methods yield more accurate predictions than unaided clinical assessments in many domains including interpersonal violence (e.g., Bonta, Law, & Hanson, 1998; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Grove & Meehl, 1996). The demonstrated accuracy of actuarial methods suggested they could be used to develop risk assessments for wife assault recidivism.

Police officers and others who work on the "front lines" of the criminal justice system must make quick decisions about detention, bail, and victim assistance. They are also in a unique

position to assess some risk factors for wife assault, not least by accessing general criminal information. They can also interview the victim about the relationship history and other risks to her safety not usually recorded in official documents. Research suggests, for example, that recent separation and the victim's fear of the offender are good predictors of repeated wife assault (Campbell, 2001; Weisz et al., 2000). There is good reason to believe, therefore, that an actuarial instrument based on information available to law enforcement officers could be used to construct an accurate risk assessment of wife assault recidivism.

The present study was designed to test the predictive validity of both information obtained by officers attending occurrences of male-to-female domestic assault, and information maintained in criminal records management systems. We also applied methods used in developing actuarial violence risk assessments to derive a "front-line" risk assessment for the prediction of wife assault recidivism. Our research questions were: 1. How accurate is the currently used risk assessment (DVSR) in predicting wife assault recidivism? 2. Can information typically available in domestic violence investigations be used to develop an actuarial risk assessment? 3. Can such a risk assessment tool be scored reliably using information typically available to officers? We also included the DA and the SARA for comparison purposes.

Method

Index Offense

From the data bases described below, and for each selected offender, we identified the occurrence closest to, but no later than, December 31, 1996, that involved a victim report or police evidence of forceful physical contact by a male against his current or former wife or common-law wife. We only considered cases in which the offender and victim had lived

together, primarily because it was difficult to distinguish noncohabiting intimate from nonintimate cohabiting relationships using the available information. Cases were considered eligible if there was evidence in the police report of both an intimate relationship and an existing or prior marital or cohabiting relationship. The offender need not have been arrested or charged for the index assault to be eligible (as the evidentiary requirements may be higher, and earlier cases appeared less likely to be charged). The first 589 offenders retrieved were selected for coding in the construction phase. A further 100 cases were reserved for cross-validation.

Data bases. The primary data came from the Ontario Municipal Provincial Police Automated Cooperative (OMPPAC) system. This electronic archive includes verbatim reports made by front line officers, including those of the Ontario Provincial Police (OPP) and approximately 50 urban police services. The OPP is the second largest police service in Canada and serves most rural areas and many municipalities in Canada=s most populous province. (Some large urban areas maintain their own separate police services and record management systems.) OMPPAC information is entered by the investigating officer and includes: names of the offender or suspect, the complainant, and the victims; charges laid; and details of the investigation.

We used the OMPPAC data to identify cases for the study by beginning with eligible occurrences (“index offenses”) dating from December 31, 1996 and working backwards through the data base, searching for all entries for each identified offender. OMPPAC was not upgraded or enhanced during the time frame of this study. We also interrogated the national Canadian Police Information Centre (CPIC) data base for all records pertaining to each identified offender. CPIC records include all criminal charges, arrests, convictions and criminal dispositions in

Canada, based on information from the Fingerprint Service of the Royal Canadian Mounted Police, a national police service. Its exhaustive record assures generalizability to similar Western industrialized nations. Both OMPPAC and CPIC are routinely available to police officers in the field, sometimes through portable computers in police vehicles, but always via terminals at local stations. We confined potential predictor variables and analyses to information available from these two sources to ensure that the resulting actuarial assessment could be completed with reasonable ease by front-line officers.

Procedure

Coding each case required research assistants to read the OMPPAC and CPIC reports and identify the index offense. Data about the index offense perpetrator, victim, relationship, and incident were coded. Then variables pertaining to the perpetrator=s sociodemographic and criminal history and outcome were coded. All variables were coded by the fourth and fifth authors who have extensive experience coding such information, or by graduate research assistants with approximately three months of training and continuous supervision. Variables describing the construction sample are shown in Table 1. The descriptive variables shown in Table 1 are mostly self-explanatory and provide a comprehensive description of the sample. For those requiring them, fuller explanations follow.

Substance Abuse Score. This score ranged from 0 to 8 with one point given for each of several items showing predictive success in previous empirical research (Harris et al., 1993): offender consumed alcohol just before or during the index offense, offender used drugs just before or during the index offense, offender abused alcohol or drugs in the few days or weeks before the index offense, offender used alcohol or drugs more than usual in the few days or

weeks before the index offense, offender is noted to be more angry or violent when he uses alcohol or drugs, offender has previously been charged for a criminal offense while under the effects of alcohol, offender had an alcohol problem since he was 18, offender had a drug problem since he was 18.

Injury to victim at index, total prior injury to partners, and total prior injury to nondomestic victims. Injury was scored on a seven point ordinal scale ranging from "1 – none" to "7 – death with mutilation."

Prior criminal history score. History was scored using the Cormier-Lang scale (Quinsey et al., 1998) which captures the frequency and severity of criminal history by totaling all offenses ranging from "1 – minor property offense" to "28 – homicide."

CTS severe violence. This variable records whether the perpetrator used acts against the partner that are defined as severe on the Conflict Tactics Scale (Straus, Hamby, Boney-McCoy, & Sugarman, 1996): used a knife or gun, punched, choked; slammed against wall; beat up; burned or scalded, kicked.

Victim barriers to support score. This variable ranged from 0 to 5 in which one point was given for each of several variables that might increase the partner's vulnerability: she has children (under age 18) living with her, she has no telephone at home, she has little or no access to a car or public transportation, she lives in an isolated location, police report that she has a substance abuse problem. Other vulnerability sources such as language barriers were also recorded but did not contribute to the predictive accuracy of this score.

We created a coding manual that captured all quantifiable information found in a preliminary review of approximately 50 cases. We chose for analysis the entire pool of items that

was coded as available in at least 2% of cases. Exceptions were the Danger Assessment (DA) and the SARA, included for comparison purposes in validating the actuarial risk assessment. They required some clinical and historical information not available to officers, so we scored these scales using Probation and Parole Office files, if the offender had served a custodial sentence or probation for any offense (55% of cases). We did include some DA or SARA items in the pool of potential predictors if they could be coded using only the OMPPAC or CPIC or data bases.

Wife Assault Recidivism

Information about subsequent criminal and assaultive behavior was coded up to the end of 2001, yielding a mean followup period of 4.79 years after the index offense (*SD* = 1.08), or 4.30 years when custodial sentences (not necessary time served) were subtracted. Any subsequent violent assault against an (ex) wife or (ex) common-law wife known to police was deemed wife assault recidivism, whether or not charges were laid. In this time frame, 175 men (30% of the 589 men in the construction sample) were recorded as having committed a subsequent assault against a female domestic partner an average of 15.1 (*SD* = 12.2) months after the index offense. In over 95% of the cases, the subsequent victim and index victim were the same person.

Analytic Strategy

First, we computed DVSR and estimated its accuracy. We then computed the bivariate relationship between each study variable and recidivism shown in Table 1. We used multivariate methods to select those variables that would produce the most efficient prediction tool, using a strategy that proved successful in developing an actuarial risk assessment in the past (Harris et al., 1993). First, we classified the potential items on a rational basis into six sets: offender=s sociodemographic characteristics, offender's domestic violence history, offender's general

criminal history, relationship characteristics, victim characteristics, and index offense details (see Table 1 for variables within each set). To minimize shrinkage on cross validation that can result from analyses capitalizing on chance, we used a simple bootstrapping procedure (Mooney & Duval, 1993). We conducted each forward conditional binary logistic regression (alpha = .05) once on each of nine randomly selected subsamples (n = 359) drawn with replacement from the full construction sample. In each set described above, only variables selected by at least five of the nine maximum likelihood estimations were retained. At the next stage, all surviving variables from the set-wise analyses were tested together in ten final forward conditional binary regression analyses (one for each of the nine subsamples and one for the full construction sample of 589). Any variable not selected in the full sample and at least one subsample was dropped. The result of this selection was named the Ontario Domestic Assault Risk Assessment (ODARA).[1] As a check for the possibility that this strategy still capitalized on chance, we tested the ODARA in the cross-validation sample of 100 new cases.

Reliability

The inter-rater reliability for the coding of the variables in Table 1 and all others in the item pool was established by having two research assistants independently score a random sub-sample of 30 cases. Only variables with Pearson correlation coefficients of at least .80 (for continuous variables) or kappa coefficients of at least .70 (for categorical variables) were retained; two (location of offense and whether the victim assaulted the offender) were rejected.

Results

As Table 1 illustrates, the wife assaulters in the construction sample were typical of those reported in the literature as generally aggressive (e.g., Holtzworth-Munroe, Meehan, Stuart,

Herron, & Rehman, 2000; Monson & Langhinrichsen-Rohling, 1998): many exhibited evidence of substance abuse, had committed prior domestic assaults, had some other form of prior criminal behavior, and were not legally married. All offenders caused at least some physical injury in the index offense, and most were charged with a criminal offense for this incident. On the other hand, few of the offenders had used a weapon, threatened serious harm or death, or attempted suicide in the index offense. Within an average opportunity time of 15 months, 29.7% of offenders recidivated but none of the recidivistic offenses involved a fatality.

Predictors of Recidivism

Table 1 shows the correlation between each variable and dichotomous wife assault recidivism. Several study variables predicted subsequent wife assault, especially those pertaining to prior criminal conduct, prior wife assault, and substance abuse. When coded from archival information and scored by summing the items present, the DVSR yielded a statistically significant correlation with recidivism in the construction sample, $r = .26$, $p < .001$. To evaluate the tradeoff between sensitivity and specificity at all possible cutpoints on the DVSR, we computed a maximum likelihood estimate of the Relative Operating Characteristic (ROC; Rice & Harris, 1995), which yielded an estimated area under the curve^2 = .67, 95% CI $\pm$ .04. Although the integrity of DA and SARA scores cannot be guaranteed because the interviews and clinical judgments recommended by their authors were not available, we also evaluated their predictive accuracies for comparison, ROC areas: DA = .59, 95% CI $\pm$ .05, SARA = .64, 95% CI $\pm$ .05.

Developing an Actuarial Risk Assessment

The 54 setwise and stepwise selection analyses yielded a mean multiple-R = .317, $p < .001$. The ten final analyses yielded a mean multiple-R = .558, $p < .001$. The construction

technique described in the Method section selected thirteen items (indicated in Table 2) together having a mean (based on Fisher's r to z transformation) Pearson product moment correlation of .21 (*SD* = .06), $p < .001$ with dichotomous recidivism. The thirteen predictor items yielded a mean Pearson product moment intercorrelation of .14 (*SD* = .12, |range| = .02-.55). To make the items easy to score, we dichotomized seven variables that were not already binary. The base rate of most variables was sufficiently low as to make "zero" versus "one or more" the appropriate dichotomization. For two variables (substance abuse history and number of children) prediction was improved by dichotomizing at "zero or one" versus "more than one" (see Table 2). After dichotomization, the thirteen items yielded a scale with possible values from 0 to 13, but with an observed range in this construction sample of 0 to 11, mean = 2.89 (*SD* = 2.14) and a Pearson (point biserial) correlation of .434, $p < .001$ with dichotomous wife assault recidivism.

The distribution of ODARA scores was positively skewed such that categorizing some of the scores (5-6, 7-13) still made for small categories (6% and 1% respectively). To evaluate the tradeoff between sensitivity and specificity at all possible cutpoints on the ODARA, we computed a maximum likelihood estimate of the area under the Relative Operating Characteristic (ROC; Rice & Harris, 1995) which is thought to be insensitive to base rate. The ODARA yielded an estimated area under the curve = .77 (*SE* = .02) 95% CI $\pm$ .04. By commonly accepted standards, this effect size is large and corresponds to a *d* of 1.1 (Cohen, 1992). An area of .77 indicates a probability of .77 that a randomly selected recidivist would have a higher score than a randomly selected nonrecidivist. Positive predictive power (PPP) ranged from .297 to .717, and negative predictive power, .703 to .957, depending on the cut point chosen. Figure 1 shows the inter-relation of all these statistics. For example, if an ODARA score of 4 were used as a cut-

point for classification as a recidivist, the sensitivity (correct classification of recidivists) = .59; the specificity (avoidance of incorrect classification of nonrecidivists) = .79; the PPP (proportion of classified recidivists who actually recidivated) = .54; and the NPP (proportion of classified nonrecidivists who do not actually recidivated) = .87. From Table 3, it can also be seen that 80% of wife assaulters in the construction sample scored 4 or lower, and 41% with this score recidivated. Finally, the high reliability and fairly low variability of the ODARA (indicated by the confidence intervals) imply low likelihood of misclassification.

Reliability of the ODARA

The OMPACC and CPIC reports used by the research assistants contained information about offenses after the index offense. Because the presence of subsequent offenses might contaminate the coding of some potential predictor variables, we arranged a stringent test of inter-rater reliability. A new research assistant separated all information for events prior to and including the index offense from all information pertaining to post-index events for a new random sub-sample of 24. This procedure required separating, photocopying, and blacking out contaminating information. Another research assistant who was not the original coder of these cases independently coded the pre-index and index information, and yet another independently coded the post-index information. These independent codings were very highly correlated[3] for both the ODARA, ICC = .90, and recidivism, ICC = .91. The ODARA inter-rater reliability yielded a standard error of measurement, SEM = .48, indicating that 95% of the time the obtained score would be expected to differ from the "true" score by $\pm$ 1.96 (.48) or less than one point. The correlation between ODARA score and recidivism in this small sample was the same whether coders were masked, r = .69, or unmasked, r = .68, yielding no evidence that unmasked

ODARA scoring was associated with better predictive accuracy.

Validation

Total score on the ODARA was correlated with the DA, $r = .43$, and the SARA, $r = .60$ and the DVSR, $r = .53$, $p < .01$, suggesting some concurrent validity with these nonactuarial assessments. In the cross-validation sample (n=100), the base rate of wife assault recidivism was slightly lower (26%) than in the construction sample; however, the ODARA yielded the same mean score of 2.89 (*SD* = 1.84). On cross-validation, its correlation with wife assault recidivism was smaller, $r = .359$, $p < .001$, corresponding to an ROC area = .72 (*SE* = .06). The DA, SARA, and DVSR all yielded statistically significant predictive validities in the construction sample, but all were significantly lower than the ODARA, as indicated by the ROC areas falling below the confidence interval for the ROC area of the ODARA. In the cross-validation sample, none of these other scales significantly predicted the outcome, ROC areas = .53, 95% CI$\pm$.14, .54, 95% CI$\pm$.14, and .59 CI$\pm$ .13, respectively. Thus, lower predictive accuracies in the cross-validation sample were more likely due to sampling error than to shrinkage.

To examine the ODARA's validity as a predictor of violence severity, we calculated the correlation between the ODARA score and several measures of the severity of each offender=s outcome in the construction sample. The measures were: the sum of victim injury scores for all subsequent domestic offenses, the sum of Cormier-Lang scale scores for all subsequent domestic offense charges, and the number of subsequent domestic incidents with acts of severe (Straus et al., 1996) violence. All measures were positively correlated with ODARA score, Pearson *r*=s = .37, .36, and .34, respectively, all *p*s < .001. ODARA score was also significantly related to the total number of subsequent occurrences of wife assault recidivism (mean = .42, *SD* = 1.14)

recorded for each perpetrator, $r = .37$, $p < .001$, and for each recidivist (mean = 2.17, *SD* = 1.70), $r = .31$, $p < .001$. Finally, ODARA score was significantly related to time at risk, $r = -.34$, $p < .001$, and among recidivists to shorter time until recidivism, $r = .26$, $p < .001$. Men who had been arrested at the index offense (50.1%) had higher ODARA scores, $r = .29$, $p < .001$.

A Field Simulation of Scoring by Police Officers

To ensure that the ODARA could be coded by police officers with the brief training likely to be available in the field, two officers not involved in the ODARA construction each independently scored the ODARA for 10 cases using data drawn from OMPPAC and CPIC data bases. The officers were provided with a first-person account of the index offense based on the narrative description in the OMPPAC record, designed to simulate a police officer's notes from a domestic violence investigation. They were also given printouts of all OMPPAC and CPIC records pertaining to the offender dated before the index offense. These are materials from which the ODARA would be scored in practice.[5] The independent police officers' scores yielded an intraclass correlation$^2 = .95$, $p < .001$.

Discussion

The methods used in the present study resulted in the development of a simple actuarial risk assessment tool, the Ontario Domestic Assault Risk Assessment, to evaluate the likelihood that a man 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.

This study also documents the power of empirical methods in developing assessments to evaluate risk of violence (see review by Monahan, 1996). Following selection of potential items

(that is, information available to police), empirical methods resulted in a risk assessment for police-known wife assaults that was not only strongly predictive but also shorter and more suited for use in law enforcement contexts than three existing nonactuarial risk assessments. Some SARA items are not easily scored by front-line law enforcement officers, and the DA was intended to assess the risk of homicide using victim interviews. The DVSR, however, was created with front-line officers in mind, using information they typically collect, and tested in this study on the population for which it was designed. Nevertheless, the DVSR did not perform as well as the ODARA, an actuarial tool derived from the same pool of information.

Despite the availability of actuarial assessments that are more accurate than clinical methods, there has been scant evidence that forensic decision makers avail themselves of these advancements in assessment tools (e.g., Hilton & Simmons, 2001; see also Harris et al., 2002; Janus & Meehl, 1997). One possible impediment is that few forensic professionals receive the training in statistical inference necessary to understand probabilistic and comparative risk (Hilton, Harris, Rawson, & Beach, in press). The present study demonstrates that a brief, easily scored, and easily interpretable actuarial assessment can yield a large prediction effect and can be scored by officers with no statistical training. We are currently evaluating the training and use of the ODARA within the Ontario Provincial Police.

The present study illustrates some pros and cons of two different approaches to the development of a formal risk assessment. One approach is exemplified by the SARA (Kropp et al., 1995) which did not directly result from a follow-up study. Rather, items were identified primarily from the literature on characteristics of assaultive husbands, the predictors of violent crime, and clinical experience. 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 has a good (but not certain) chance of yielding a correlation with wife assault recidivism. Because a sum of scores represents a formulaic or mechanical system, it is likely to perform better than unaided clinical judgment (Grove & Meehl, 1996) as long as some items are valid predictors. Because the items are selected from the entire literature, generalizability may be less problematic than an instrument constructed using one particular population. On the other hand, subsequent empirical validation is required before a nonactuarial approach can provide data on reliability, accuracy, and population norms for prediction; such data remain limited for the structured clinical assessments described here (e.g., Grann & Wedin, 2002).

Conversely, the actuarial method, exemplified by the ODARA, is more clearly based on established psychological assessment techniques. Evidence of validity and reliability, as well as norms and specific prediction values, is inherent in the construction of an actuarial assessment. Actuarial methods are also formulaic and likely to outperform unaided clinical judgment. On the other hand, even with cross-validation, the development of an actuarial method can usually be based only on a single population. Additional research is then required to firmly establish generalization to other relevant populations. Fortunately, the predictors of violent crime appear to be extremely general (Bonta et al., 1998; Quinsey et al., 1998) and actuarial instruments for violence prediction exhibit high generalizability (see review in Harris & Rice, 2003).

An important advantage of actuarial assessment for law enforcement services and the courts they serve is that each score corresponds to a percentile rank in the referent population and an estimated probability of the outcome. Using actuarial scores, for example, a decision could be made to deny bail to offenders when (among other appropriate criteria) the probability of

recidivism within a particular time frame was above, say, 60%, or to the most dangerous 20% of offenders. Similarly, judgments about cut points for forensic decisions can be made on an individual or policy basis, informed by norms showing the proportion of offenders affected and the estimated number of assaults prevented. In contrast, after scoring a typical nonactuarial assessment, assessors must then make an unaided clinical judgment about an offender's relative and absolute risk, and subsequent risk management decisions must be made without knowledge of either the proportion of cases affected or the probable reduction of risk.

In our jurisdiction, we anticipate that the ODARA would be scored by officers immediately after completing a domestic violence investigation. Thus, scoring would not be done amid the conflict and danger of the domestic violence situation but would nevertheless be based on all the information usually gathered by the officers prior to a bail hearing. Because a bail hearing usually occurs within hours of an arrest, a brief assessment is essential. In bail and other conditional release decisions, it is not culpability but risk of another offense (and other failures to comply with conditions) that is at issue. In the present study, arrest decisions were positively, but far from perfectly, correlated with ODARA score, indicating that knowledge of the likelihood of recidivism, along with an actuarial measure of the accused’s rank among similar men, could aid such decisions.

There were limitations in the present study. First, the index offense could not always be identified without reading all the recorded material, and the resources required to review, identify, separate, and independently code predictor and outcome variables were so prohibitive that we limited this effort to a randomly selected subgroup. The masked coding of ODARA and outcome, however, yielded excellent interrater reliabilities and no evidence that the predictive

association between the ODARA and recidivism was inflated by unmasked coding. Prospective replication studies using entirely blind follow-up are currently underway.

Second, no lethal assaults were committed by offenders in this study. The Danger Assessment (Campbell, 1986) was designed to predict homicide and, therefore, was included here only to permit a comparison of its variation on replication with the shrinkage in cross-validation on the ODARA. We make no claim as to the superiority of the ODARA in predicting lethal recidivism. ODARA scores were, nevertheless, associated with the severity of injuries caused in recidivism. Prospective prediction of domestic murder is hampered by extremely low base rates and by the large proportion of murderers who have no prior criminal history (Eke, Hilton, & Harris, 2003).[6] We argue that the occurrence of nonlethal domestic assault is itself a serious criminal justice and public health concern. As well, it can be assumed that lethal assaults can only result from relatively serious injuries.

Third, we did not include measures of psychopathy and other psychological variables known to be strong predictors of violent recidivism by severe wife assaulters (Grann & Wedin, 2002; Hilton et al., 2001) and of violent reoffending in general (Quinsey et al., 1998). Our aim was to develop an accurate risk assessment suitable for front-line law enforcement officers, who have neither the time nor the training to score the Psychopathy Checklist Revised (PCL-R; Hare,1991) and similar psychological assessments. We are currently examining whether the ODARA can be improved by adding PCL-R score, childhood exposure to violence, and other information typically available to those who work with offenders over a longer duration, such as correctional staff, probation and parole officers, and mental health professionals. This research may determine whether predictors of general violence perform equally well in predicting

recidivism among less severe wife assaulters. Although some ODARA items were dichotomized to make the assessment easier for front-line officers to use, dichotomizing did not significantly lower predictive power. We will examine the benefit of keeping continuous variables for the in-depth assessment we are currently developing.

Finally, although some of the cases we studied included spousal assault by women in addition to men's violence, the ODARA is not yet adapted for this outcome. Furthermore, although many women are assaulted or murdered by their boyfriends or ex-boyfriends, we excluded non-cohabiting dating couples. We cannot yet recommend the ODARA for predicting violence for these groups until validation studies are conducted. Because over 95% of the recidivists in our sample reoffended against the same spouse, a woman's risk of re-victimization can be estimated by her partner's ODARA score. We are currently testing a clinic format of the ODARA suitable for victim interview, and studying the predictive validity of additional victim barriers and supports (e.g., access to income).

In summary, 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 on both construction and cross-validation samples. This study suggests a substantial increase in predictive accuracy could be gained by policing services by simply replacing current rational risk assessments with a validated actuarial method. In our view, the present study also illustrates the value of the methods from psychological assessment research in the development of decision tools for use by nonpsychologists in many fields, especially criminal justice and forensic decision-making.

References

Aldarondo, E., & Sugarman, D.B. (1996). Risk marker analysis of the cessation and persistence of wife assault. *Journal of Consulting and Clinical Psychology, 64,* 1010-1019.

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

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

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

Campbell, J. (2001, June). *Risk assessment for intimate partner femicide. What practitioners need to know.* Presented at the International Conference on Children Exposed to Domestic Violence. London, ON.

Campbell, J.C., Soeken, K.K., McFarlane, J., & Parker, B. (1998). Risk factors for femicide among pregnant and nonpregnant battered women. In J.C. Campbell (Ed.), *Empowering survivors of abuse: Health care for battered women and their children,* (pp. 90-97). Thousand Oaks, CA: Sage.

Cohen, J.A. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

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

Dutton, D.G., Bodnarchuk, M., Kropp, R., Hart, S.D., & Ogloff, J.R. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. *International Journal of Offender Therapy and Comparative Criminology, 41*, 9-23.

Eke, A.W., Hilton, N.Z., & Harris, G.T. (2003). Intimate partner murder. Manuscript in preparation.

Glover, A.J.J., Nicholson, D.E., Hemmati, T., Bernfeld, G.A., & Quinsey, V.L. (2002). A comparison of predictors of general and violent recidivism among high risk federal offenders. *Criminal Justice and Behavior, 29*, 235-249.

Goodman, L.A., Dutton, M.A., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers. *Journal of Interpersonal Violence, 15*, 63-74.

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

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

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

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior,* 24, 119-136.

Hare, R.D. (1991). *The Hare Psychopathy Checklist-Revised. Manual.* Toronto: Multi-Health Systems.

Harris, G.T., & Rice, M.E. (2003). Actuarial assessment of risk among sex offenders. In R.A.

Prentky, E. Janus, & M.C. Seto (Eds.). *Understanding and managing sexually coercive behaviour* (pp. 198-210). New York: Annals of the New York Academy of Sciences, Vol. 989.

Harris, G.T., Rice, M.E., & Camilleri, J.A. (in press). Applying a forensic actuarial assessment (the *Violence Risk Appraisal Guide*) to nonforensic patients. *Journal of Interpersonal Violence.*

Harris, G.T., Rice, M.E., & Cormier, C.A. (2002). Prospective replication of the *Violence Risk Appraisal Guide* in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26,* 377-394.

Harris, G.T., Rice, M.E., & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315-335.

Harris, G.T., Rice, M.E., Quinsey, V.L., LalumiПre, M.L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15,* 413-425.

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

Hilton, N.Z., Harris, G.T., Rawson, K., & Beach, C. (in press). Communication of risk information to forensic decision-makers. *Criminal Justice and Behavior.*

Hilton, N.Z., & Simmons, J. L. (2001). Actuarial and clinical risk assessment in decisions to release mentally disordered offenders from maximum security. *Law and Human Behavior, 25,* 393-408.

Holtzworth-Munroe, A., Meehan, J.C., Stuart, G.L., Herron, K., & Rehman, U. (2000). Testing

the Holtzworth-Munroe and Stuart (1994) batterer typology. *Journal of Consulting and Clinical Psychology, 68,* 1000-1019.

Janus, E.S., & Meehl, P.E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy, and Law, 3,* 33-64.

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

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

McGraw, K.O. & Wong, S.P. (1992). A common language effect size statistic. *Psychological Bulletin, 111,* 361-365.

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

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

Monson, C.M., & Langhinrichsen-Rohling, J. (1998). Sexual and nonsexual marital aggression: Legal considerations, epidemiology, and an integrated typology of perpetrators. *Aggression and Violent Behavior, 3,* 369-389.

Mooney, C.Z., & Duval, R.D. (1995). *Bootstraping: A nonparametric approach to statistical inference.* Newbury Park, CA: Sage.

Office of the Chief Coroner of Ontario (1998, July). *Jury's Verdict and Recommendations from the Inquest into the Deaths of Arlene May and Randy Iles.* Toronto, Ontario: Coroners Courts. Available on line at: http://www.owjn.org/archive/arlene3.htm.

Office of the Chief Coroner of Ontario (2002, February). *Jury's Verdict and Recommendations from the Inquest into the Deaths of Gillian and Ralph Hadley.* Toronto, Ontario: Coroners Courts. Available on line at: http://www.owjn.org/issues/w-abuse/hadley2.htm.

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

Rice, M.E. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist, 52,* 414-423.

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

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

Rice, M.E., & Harris, G.T. (2002). Men who molest their sexually immature daughters: Is a special explanation required? *Journal of Abnormal Psychology, 111,* 329-339.

Roehl, J., & Guertin, K. (2000). Intimate partner violence: The current use of risk assessments in sentencing orders. *The Justice System Journal, 21,* 171-198.

Saunders, D.G. (1993). Husbands who assault: Multiple profiles requiring multiple responses. In N. Z. Hilton (Ed.), *Legal responses to wife assault* (pp. 9-34). Newbury Park, CA: Sage.

Shepard, M. (1992). Predicting batterer recidivism five years after community intervention.

Journal of Family Violence, 7, 167-178.

Straus, M.A., Hamby, S.L., Boney-McCoy, S., & Sugarman, D.B. (1996). The Revised Conflict Tactics Scale (CTS2). *Journal of Family Issues, 17,* 283-316.

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

Weisz, A.N., Tolman, R.M., & Saunders, D.G. (2000). Assessing the risk of severe domestic violence. *Journal of Interpersonal Violence, 15,* 75-90.

Author Note

We would like to thank the Ontario Ministry of Public Safety and Security (Ministry of Corrections) for permission to access information and for assistance in accessing files; Ruth Houghton, Joseph Camilleri, Leslie Belchamber, Sonja Dey, Angela W. Eke, Marnie Foster, Kelly Grubb, Tracy Lowe-Wetmore, Julie McKay, Kelly Rawson, Amilynn Bryans, Franco Bassiorello, and Steve Hayward for research and administrative assistance. Funding for this research was provided by the Social Sciences and Humanities Research Council. The opinions expressed are the authors'. Correspondence may be addressed to N. Zoe Hilton, Ph.D., at zhilton@mhcp.on.ca, or by mail to Research Department, Mental Health Centre, Penetanguishene, ON, L9M 1G3, Canada. Email:

Footnotes

[1] In order to check that our setwise method had not unduly influenced the variables selected for the instrument, we conducted a logistic regression on the entire set of all potential predictors. This analysis yielded an identical R^2 and selected 13 variables, including 10 identical to those selected by our original procedure. The three different variables included a criminal history item (total criminal history score instead of conditional release violations), and domestic violence and relationship items (number of prior assaults against domestic victims and any relationship separations in the year before the index, instead of threats and confinement in the index offense).

[2]ROC area is conceptually and numerically equal to the Common Language Effect size (McGraw & Wong, 1992; Rice & Harris, 1995).

[3]Six computation models for the intra-class correlation coefficient yielded the same values.

[4]The ROC area for the scale with undichotomized items was larger but not statistically significantly so, area = .80, 95% CI $\pm$ .04..

[5]The simulated field trial materials are available (anonymized) from the first author upon request.

[6]In a separate sample, of the next 600 identified cases, one man committed a lethal wife assault recidivism. His ODARA score at index was 8; i.e., in the top 1% of known wife assaulters.

Table 1

Sample Characteristics and Correlation with Wife Assault Recidivism

	Mean (SD) or %	Correlation with Recidivism
Offender's Sociodemographic Characteristics		
Unemployed (%)	20	.15**
Appeared suicidal (%)	6	.06
Age	38.2 (12.0)	-.14**
Substance abuse score (0-8)[a]	1.31 (1.47)	.29***
Offender's Domestic Violence History		
Ever violated prior no contact order (%)	5	.11**
Ever assaulted victim when pregnant (%)	3	.13**
Prior domestic incidents (OMPPAC)	.39 (90)	.29**
Total prior injury to partner(s) (1-7)[a]	1.19 (.82)	.21***
Danger Assessment (0-15)	.48 (.99)	.20***
SARA (0-40)	3.11 (4.14)	.27***
DVSR (0-22)	1.40 (1.50)	.26***
Offender's General Criminal History		
Prior correctional sentence (%)	24	.28***

(*Table continues*)

	Mean (SD) or %	Correlation with Recidivism
Offender violent towards others (%)	4	.20***
Any prior violation of conditional release (%)	15	.25***
Number prior criminal charges	3.30 (5.82)	.24***
Prior criminal history score[a]	5.40 (11.4)	.17***
Prior nondomestic incidents (OMPPAC)	.09 (.40)	.17***
Total prior injury to nondomestic victims[a]	1.09 (.77)	.11**
Relationship Characteristics		
Reported sexual jealousy (%)	7	.12**
Separation prior to index (%)	28	.01
Not legally married (%)	57	.11**
Duration of relationship (mo)	93.4 (107)	-.16***
Total number of children	1.85 (1.52)	.26***
Victim Characteristics		
Victim Unemployed (%)	21	.20***
Victim Age	34.6 (11.1)	-.14**
Number of children from prior relationships	.45 (.84)	.23***
Barriers to support score (0-5)	.81 (.72)	.21***

(*Table continues*)

	Means (SD) or %	Correlation with Recidivism
Victim reports offender is violent outside the home (%)	4	.20***
Offense Details		
Alcohol involved (%)	43	.11**
Offender threatened harm/death (%)	15	.12**
Weapon involved (%)	9	.03
Victim feared future violence (%)	10	.14**
Offender confined victim (%)	7	.12**
Perpetrator charged (%)	53	.15**
CTS severe violence (%)[a]	38	.06
Index location was shared home (%)	75	.04
Mutual assault (%)	32	-.07
Victim's injury score (1-7)	2.03 (.96)	.06

Note: Correlations are Pearson product moment correlations (i.e., phi for dichotomous variables, point-biserial for continuous measures). [a]Variables described in text. *$p < .05$, **$p < .01$, ***$p < .001$.

Table 2

Dichotomous Variables Used in the Ontario Domestic Assault Risk Assessment (ODARA). Also Shown is the Phi-Coefficient for Each Item and the Dichotomous Recidivism (all $p < .01$).

Variable	Correlation with Recidivism (phi)
1. Has a prior domestic assault (against a partner or child) in police RMS	.26
2. Has a prior nondomestic assault (against anyone other than a partner or child) in RMS	.15
3. Has a prior sentence to a term of 30 days or more	.28
4. Has a prior failure on conditional release including bail, parole, probation, no-contact order	.25
5. Threatened to harm or kill anyone during index offense	.12
6. Unlawful confinement of victim during index offense	.12
7. Victim fears repetition of violence	.14
8. Victim and/or offender have more than one child altogether	.24
9. Offender is in step-father role in this relationship	.22
10. Offender is violent outside the home (to people other than a partner or child)	.20
11. Offender has more than one indicator of substance abuse problem	.27
12. Offender has ever assaulted victim when she was pregnant	.13
13. Victim faces at least one barrier to support	.11

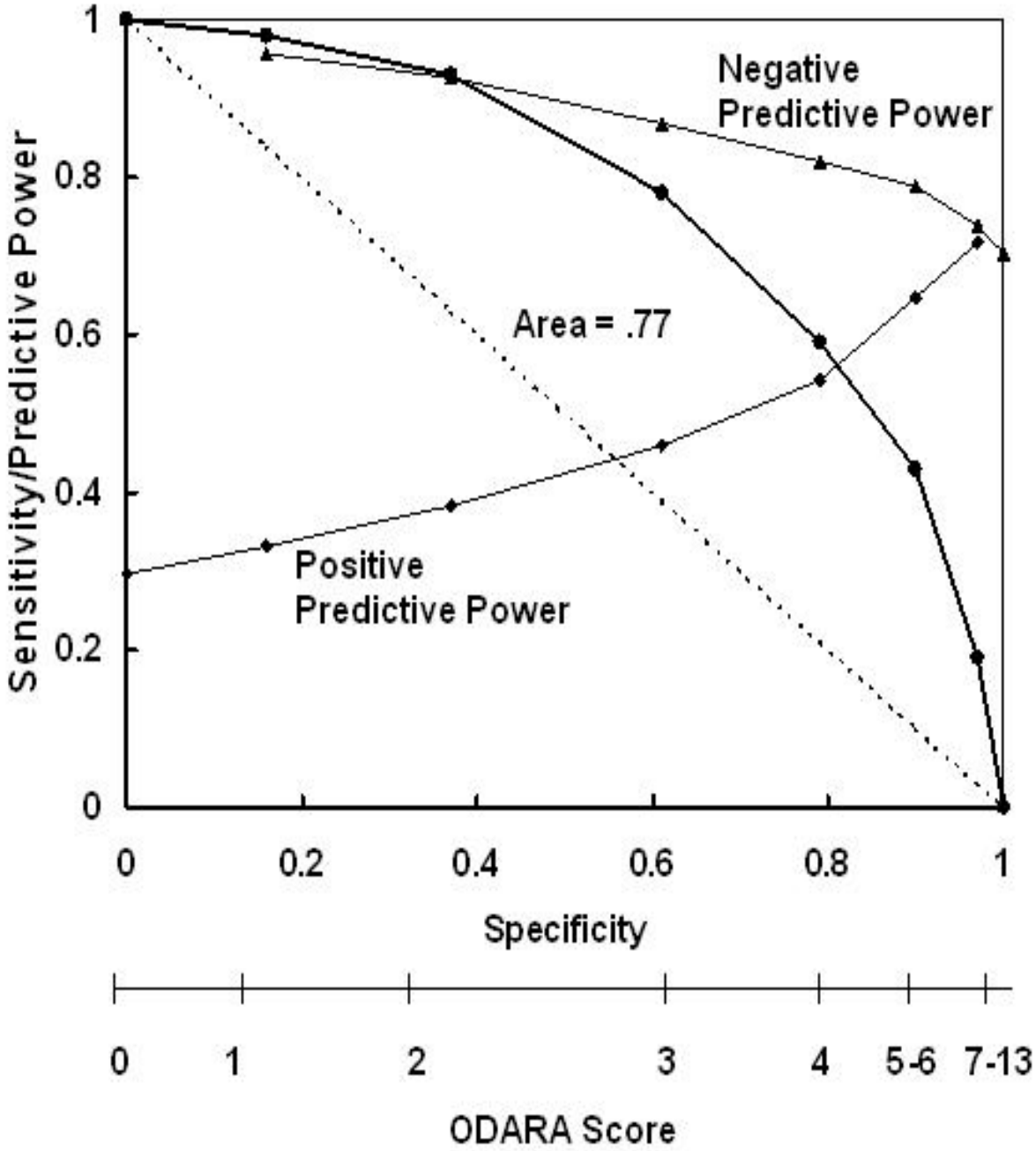
Table 3

Interpretation of Scores on the Ontario Domestic Assault Risk Assessment

Score	Category	Cumulative Proportion	Overall Recidivism Rate	95%CI
0	1	11	.05	±.049
1	2	27	.10	±.056
2	3	48	.20	±.065
3	4	67	.27	±.076
4	5	80	.41	±.102
5 or 6	6	93	.59	±.102
7 - 13	7	99	.70	±.130

Figure Caption

Figure 1. Relative Operating Characteristic and other accuracy statistics for ODARA predictions of dichotomous wife assault recidivism; sensitivity (circles) as a function of specificity on the upper abscissa; and Positive Predictive Power (PPP, diamonds), and Negative Predictive Power (NPP, triangles) as a function of ODARA score on the lower abscissa.


Sensitivity/Predictive Power
1
0.8
0.6
0.4
0.2
0
Negative
Predictive Power
Area = .77
Positive
Predictive Power
0
0.2
0.4
0.6
0.8
1
Specificity
0
1
2
3
4
5-6
7-13
ODARA Score