

Battered Women's Perceptions of Risk Versus Risk Factors and Instruments in Predicting Repeat Reassault

D. ALEX HECKERT

Indiana University of Pennsylvania

EDWARD W. GONDOLF

*Mid-Atlantic Addiction Training Institute
and Indiana University of Pennsylvania*

This study partially replicates and expands on a previous study that showed women's perceptions of risk to be a strong predictor of reassault among batterers. The current study employed a larger and multisite sample, a longer follow-up period of 15 months, and multiple outcomes including "repeated reassault" (n = 499). According to the multinomial logistic regressions, women's perceptions of risk improved prediction with risk factors (ROC area under the curve improved by .04 and sensitivity of repeated reassault increased 12 percentage points). In comparison to simulated risk instruments, women's perceptions by themselves were better predictors than the K-SID, similar in predictive ability to the SARA, and almost as strong as the DAS. The best prediction of repeated reassault was obtained using risk markers, including women's perceptions (ROC AUC = .83; 70% sensitivity) or by using the DAS and women's perceptions together (ROC AUC = .73; 64% sensitivity).

Keywords: *batterer; batterer programs; perceptions of risk; risk assessment; risk instruments*

The effort to identify and contain the most dangerous men has helped to promote a surge of prediction research in the domestic violence field. The two

Authors' Note: The authors wish to thank the program directors at the four research sites, who offered essential assistance in developing and implementing the data collection, and staff at Mid-Atlantic Addiction Training Institute who provided a variety of supports. They would also like to thank Dr. Alison Snow Jones for advice on data analysis. Data coding, management, and analysis was supported by the National Institute of Justice, U.S. Department of Justice (Grant No. 98-WT-VX-0014). Data collection and data management was supported through grants from the Centers for Disease Control and Prevention (CDC), U.S. Department of Health and

JOURNAL OF INTERPERSONAL VIOLENCE, Vol. 19 No. 7, July 2004 778-800

DOI: 10.1177/0886260504265619

© 2004 Sage Publications

main and related research approaches have been, one, to identify risk factors associated with reassault, and, two, to develop risk assessment instruments. The risk factors and the instrument items are primarily characteristics of the male perpetrator assessed at intake into a batterer program. Overall prediction using both approaches is relatively weak, however, with a modest ability to accurately classify reassault much beyond chance (Goodman, Dutton, & Bennett, 2000; Hanson & Wallace-Capretta, 2000; Hilton, Harris, & Rice, 2001; Limandri & Sheridan, 1995; Saunders, 1995; Weisz, Tolman, & Saunders, 2000).

There is some preliminary evidence, however, that women's perceptions of the likelihood of reassault may substantially improve prediction and warrant further consideration in our prediction efforts. A recent study comparing women's predictions with items from one risk assessment instrument, the Danger Assessment Scale (DAS; Campbell, 1986, 1995), found the women's perceptions were better predictors than the DAS items or score (Weisz et al., 2000). The small sample size ($n = 177$) and limited follow-up of 4 months point to the need for additional research. In the current study, we use a comprehensive database with a large, multisite sample of batterers and a 15-month follow-up to assess whether women's perceptions of risk improve prediction of repeat reassault above and beyond other risk factors. In addition, we compare the predictive ability of women's perceptions of risk to simulated versions of three popular risk assessment instruments.

Previous Research

A long line of research attempts to identify risk factors that might help predict continuing and escalating violence. Some fairly consistent risk factors have been identified, such as prior assault, excessive alcohol or drug abuse, previous criminality, severe personality disorders and/or psychological problems, neglect or abuse as a child, and program dropout (see DeMaris & Jackson, 1987; Dutton et al., 1997; Gondolf, 1997; Hamberger & Hastings, 1990; Hilton et al., 2001; Saunders, 1995; Tolman & Bennett, 1990). Nevertheless, the predictive power of these factors, even when combined together, is arguably weak (Limandri & Sheridan, 1995; Saunders, 1995). Many men who do not reassault are predicted to be "reassaulters" or at "high risk" for reassault (false positives), and many men who do reassault are

Human Services (Grant No. R49/CCR310525-02). The conclusions do not necessarily represent the official view of the CDC or the NIH.

predicted to be “non-reassaulters” or at “low risk” for reassault (false negatives).

To improve prediction and the identification of men likely to reassault, researchers have developed several risk assessment instruments, such as the DAS, the Kingston Screening Instrument for Domestic Violence Offenders (K-SID), and the SARA (Spousal Assault Risk Assessment). These instruments combine established risk factors under some theoretical assumptions about violence. They may rely on self-administration or clinical ratings. Their correct classification of continued violence improves on clinical judgment alone but is still limited (Kropp & Hart, 2000). Recent reviews of such instruments have, consequently, urged that such instruments be used with caution and with a variety of other sources, along with women’s predictions and safety concerns (Dutton & Kropp, 2000; Roehl & Guertin, 2000).

Debate has emerged as to whether specialized risk instruments to predict domestic violence are necessary. Hanson and Wallace-Capretta (2000), for example, found that a structured risk measure designed for general offenders, the Level of Service Inventory–Revised (LSI-R), and age were the only multivariate predictors of violence recidivism (as measured by arrests) in a Canadian sample. The LSI-R is based on social learning theory and includes subscales pertaining to criminal history, problems with education and employment, finances, relationships (including attitudes about women, partner assault, and service providers), unstable housing patterns, leisure activities, and substance abuse. Because most of this violence recidivism was domestic violence, Hanson and Wallace-Capretta have contended that the LSI-R should be used as a risk instrument for domestic violence unless future research demonstrates the superiority of instruments developed specifically for predicting domestic violence. Nevertheless, the authors were not able to compare the predictive ability of the LSI-R to a specialized domestic violence instrument using the same sample.

A recent study (Hilton et al., 2001) found that the Violence Risk Appraisal Guide (VRAG) was a slightly better, although modest, predictor of violence recidivism (not just partner reassault) among 88 serious wife assaulters ($r = .42$) than the Hare Psychopathy Checklist–Revised ($r = .35$). The VRAG is a 12-item risk assessment instrument that has been moderately successful in predicting violent recidivism among violent offenders (ROC area under the curve = .74 at 10-year follow-up) (Rice & Harris, 1995). It includes static demographic variables, childhood history, criminal history, and psychiatric assessment variables. Like Hanson and Wallace-Capretta (2000), Hilton and her colleagues (2001) questioned whether a specialized risk assessment instrument is necessary for predicting future violence among wife assaulters.

Nevertheless, their results are limited by the small sample size ($n = 88$) and the extreme and homogeneous nature of their sample of wife assaulters. All the subjects had been admitted to a maximum-security psychiatric facility and more than half (59%) had an index offense of murder or manslaughter.

Even though battered women's advocates have long recommended listening and responding to women's fears and concerns, little research has substantiated the predictive validity of these "perceptions." In the one major study, Weisz and her colleagues (2000) analyzed data on 177 women who were partners of men charged with misdemeanors and found guilty, offered probation before judgment, or given deferred prosecution. Some of the men were ordered into a batterer program, whereas others were part of a no-treatment comparison group. The partners of these batterers were asked on a scale of 0 to 10 how likely the batterer was to become violent during a dispute in the next year. This item was the single best predictor of severe violence, as measured by an expanded version of the Conflict Tactics Scale (CTS; Straus, 1979). This study showed that the women's perception of risk significantly improved prediction of severe reassault over 4 months after controlling for more than 20 other risk factors (R^2 increase = .10), and this perception of risk significantly improved prediction above and beyond 12 of the items used in Campbell's DAS (R^2 increase = .13). Because of the small sample and limited follow-up of this study, however, there is a need to extend this research using larger, more comprehensive databases.

There is another reason for extending the research by Weisz and her colleagues (2000). The previous predictive research may have produced relatively weak predictions because of dichotomized "reassault versus no reassault" outcomes. Research in the area of general violence suggests that prediction is likely to be improved with multiple outcomes that include especially a category of the outcome of most concern—severe and repetitive violence (Mulvey & Lidz, 1993; Steadman et al., 1994). Our own research using multinomial logistic regressions shows an improvement in prediction using multiple outcomes and also suggests the importance of predicting repeat reassault as opposed to just any reassault (Heckert & Gondolf, 2003). The multiple outcomes included no abuse, verbal abuse or controlling behavior, threats, one-time reassault, and repeated reassaults. The multiple outcome model successfully predicted 70% of the "repeat reassaulters," whereas the dichotomous outcome model (any reassault versus no reassault) only predicted 44% of "any reassaulters." These findings suggest the importance of assessing how well women's perceptions of risk predict multiple outcomes and especially repeat reassault.

METHOD

To test the predictive power of women's perceptions of risk, we used a database of batterers and their female partners from a multisite evaluation of batterer intervention systems. This database includes 840 men who were admitted to batterer programs in four cities—Pittsburgh, Dallas, Houston, and Denver. Interviews were conducted with batterers, initial victims, and new female partners at program intake and every 3 months over a 15-month follow-up. Using multinomial logistic regression, we first built a prediction equation with conventional risk factors gathered from self-reports of the batterers and their partners. We then added women's characteristics, including women's perceptions of risk, to determine whether prediction of repeat reassault improved. The comprehensive nature of the data set allowed us to consider most potential risk factors suggested by theory and previous prediction research. Second, we ran separate multinomial logistic regressions with the multiple outcomes variable on the three simulated risk instruments and on the women's perceptions of risk.

Research Sites

The database offers a large representative sample of batterers across four sites and diverse regions. The counseling modalities of the four batterer programs conform to the parameters of the prevailing state standards, which endorse cognitive-behavioral techniques taught in a group setting. The programs were selected to represent a range of format in terms of program duration and additional services. The format ranged from 3 months in duration and few additional services in Pittsburgh to 9 months in duration and comprehensive services in Denver (e.g., in-house alcohol treatment, individual psychological counseling, and women's service coordination).

Sample

The first 20 to 25 men appearing for program intake each month were recruited into the sample until approximately 210 men had been recruited at each of the four research sites. The target of 210 men at each site was based on a power analysis for the initial four-site comparison. The refusal rate was extremely low, at less than 5% at each of the four research sites, suggesting a representative sampling of the program referrals for 1995. There were no significant differences in the characteristics of the refusers and the participating sample. The vast majority of the men (82%) were mandated to the programs by the courts, as opposed to others who entered the programs voluntarily

(18%). The demographics of the batterers suggest a sample typical of batterers in court-referred batterer programs, although the batterers in this sample are more racially diverse, less likely to be married, and more likely to be unemployed than batterers described in studies conducted in the Midwest (Hamberger & Hastings, 1988; Saunders, 1996) and Canada (Dutton, 1986). (See Gondolf, 1999, for additional description of the study design, sample recruitment, and sample demographics.)

Data Collection

At program intake, trained research assistants administered a background questionnaire to the men that included questions about the incident that led to batterer program referral: a series of open-ended questions, followed by the CTS items for "physical aggression" (Straus, 1979). An alcohol screening test, personality inventory, and forms soliciting written consent to participate in the program evaluation were also administered. The background questionnaire additionally asked about the men's demographics, living situation, parents' behavior, mental health problems, alcohol use, prior treatment and counseling, abusive behavior, previous arrests, partner's response, and partner's help seeking.

The men's partners were interviewed by phone within 2 weeks of the men's program intake. The women were administered a background questionnaire to verify the men's reports of abuse, alcohol abuse, mental health symptoms, previous violence, the women's response to the abuse, their help seeking, their own alcohol and drug use, and other assistance or treatment the women may have received. Women respondents were contacted for 82% of the men ($n = 688$) at program intake. Although we were unable to contact the partners for the remaining 18% of the men, the women respondents appear to be representative of the base sample of cases (see Gondolf, 1999).

The men and their female partners were called separately every 3 months for a 15-month follow-up period and interviewed about their relationship status, the men's behavior toward their partner, the men's alcohol and drug use, and other treatment and assistance that the men or their partners may have received. Approximately half the men reported at program intake that they were no longer living with the initial victim. They were asked whether they had a new partner and how often they saw the initial partner. The initial victims also identified a small number of the new partners (see Gondolf, 1997).

A female partner was interviewed for 79% ($n = 662$) of the batterers at least once during the 15-month follow-up. We accounted for 67% of the women for the full 15-month follow-up. A new female partner was interviewed for 113 (14%) of the batterers, and a new partner and the initial victim

were interviewed for 68 (8%) of the men. A sophisticated attrition analysis shows negligible bias resulting from sample attrition with regard to multivariate analyses of reassault (Jones, 2002).

Measures

Abuse and reassault outcome. The multiple outcome variable was constructed from the cumulative follow-up data. Assaults and other forms of abuse were assessed through a series of questions that included the following: (a) an open-ended question about "how was the relationship going," (b) descriptions of any conflicts and their circumstances, and (c) inventories for controlling behavior, verbal abuse, threats, and physical aggression. This funneling form of questioning is recommended to increase rapport and increase self-disclosure by allowing the interviewee to tell their "story" (Spradley, 1979). The items for physical aggression were taken from the CTS (Straus, 1979); and the other forms of abuse were drawn from the CTS and two other inventories with non-physical-abuse items (Shepard & Campbell, 1992; Tolman, 1989). Women's reports of assault were used because of substantial underreporting by men during the follow-ups (Heckert & Gondolf, 2000a, 2000b).

Women's outcomes were coded if they completed at least the 9-month follow-up interview, even if the 12-month and 15-month interviews were missing. This decision was made to preserve as much of the sample as possible and is justified by the following. One, the majority of reassaults occurred within the first 6 months; the amount of physical reassault and psychological abuse steadily declined over time (Gondolf, 1997; Gondolf, Heckert, & Kimmel, 2002). Two, we conducted an analysis with only the 348 women who had complete follow-up data for all five follow-ups, in which we calculated their multiple outcome based on the first three follow-ups and cross-tabulated it with their multiple outcome based on all five follow-ups. The Kappa for this cross-classification was a robust .75, justifying the use of incomplete cases (Heckert & Gondolf, 2002).

For the multiple outcome of abuse and reassault, men were classified into the five following outcomes based on their partners' reports of abuse during the 15-month follow-up: *repeat reassaulters* with more than one incident that included one of the tactics on the physical aggression subscale of the CTS (23%); *one-time reassaulters* with only one incident of physical aggression (12%); *threatening reassault* with no physical tactics but a threat of some kind (i.e., to hit, attack, or harm; to kill; to take away children or harm them, to kill or seriously harm other people; to kill or hurt himself) (20%); *control-*

ling behavior or verbal abuse with no threats or physical tactics but any controlling behaviors or verbal abuse (i.e., kept from talking on phone; kept from friends; stopped from going someplace; followed partner; kept from using family income; took or stole money from partner; swore or screamed; accused partner of being with another man; insulted or put down; threw, smashed, hit, or kicked something; destroyed property; or hurt a pet or pets) (26%); and *no abuse* with no reports of physical assault, threats, controlling behavior, or verbal abuse over 15 months (19%).

Potential Risk Factors

Social background variables. The potential risk factors and control variables were drawn from the men's and women's background questionnaires administered at program intake. This selection of variables approximates those variables considered in previous prediction research and reflects prevailing theory (see Saunders, 1995; Tolman & Bennett, 1990). Social background variables such as age, race, employment status, and socioeconomic status (i.e., occupational status, education, and public assistance) were used. Relationship characteristics included marital status, living situation (i.e., living apart from the victim versus living together), duration of the relationship, and children living with the man. (For more details on the coding of the variables, see the appendix of Heckert & Gondolf, 2002.)

Batterer behavior. Several measures of the man's alcohol and drug use and contact with systems of social control were used. Different alcohol measures were tried because of the tendency for discrepancies between alcohol tests and self-reports and the strong influence of alcohol variables in previous prediction research (see Saunders, 1995; Tolman & Bennett, 1990). Measures of alcohol use were the following: an elevated Michigan Alcohol Screening Test (MAST; Selzer, 1971) score (> 4), alcohol-related crime reported on the MAST, man's report of being frequently drunk or high in the last year adjusted by the woman's report, and man's report of alcohol and/or drug treatment. Arrested for anything other than domestic violence in the previous year was also computed from the men's and women's reports combined.

The men's previous abuse and reassault was assessed using the same procedures as for the outcomes, except the time frames "ever in the past" and "the previous 3 months" were used. "Severe violence" was calculated instead of "repeat violence." It was considered a positive response to any of the so-called "severe items" on the CTS (hit with a fist, bit, kicked; hit or tried to hit

with something; choked or burned; threatened with or used a weapon; forced sex). The impact of past violence was measured using the women's reports of injuries, medical assistance, and hospitalization. Separate measures of controlling behavior, verbal abuse, and threats were also used.

Batterer personality. In response to the increased attention to the influence of batterer personality on outcomes, a series of variables were constructed from the Millon Clinical Multiaxial Inventory, Version III (MCMI-III; Millon, 1994). The MCMI is a 175-item test with 24 subscales that correspond to Axis II (personality disorders: e.g., antisocial or narcissistic personality) and Axis I (major disorders: e.g., major depression, schizophrenia) diagnoses. We considered individual elevated subscales as well as elevations for a category of subscales suggesting "severe" pathology, as defined by the test developers.

We also developed measures of primary psychopathy, secondary psychopathy, and psychopathic tendencies from the MCMI-III profiles using procedures recommended by experts on psychopathy and the MCMI (Blackburn, 1998; Millon & Davis, 1998). (For a full discussion of these procedures, justifications and outcomes, see Gondolf & White, 2001.)

Batterer personality types were also determined through a cluster analysis of the MCMI-III data (Gondolf, 2003). Four groups emerged from the cluster analysis: (a) little or no pathology, (b) narcissism or antisocial, (c) dependent/avoidant, and (d) multiple disorders. These types were established following the procedures used in previous studies with the MCMI and batterers (Hamberger, Lohr, Bonge, & Tolin, 1996) and approximate the prevailing batterer typologies (see Holtzworth-Munroe & Stuart, 1994).

Women's characteristics. Women's perceptions of risk, previous help-seeking behavior, and service contact were obtained through the interviews conducted with women at the time of program intake. To assess women's "perceptions of risk", the women were asked the following questions with Likert-type response choices, "How safe do you feel at this point?" and "How likely is it that your husband will become violent towards you during the next 3 months?" The responses were collapsed from their original 5-point scales based on conceptual reasons and the number of responses in each category. The safety question responses were collapsed into 1 = *uncertain, not safe, in much danger*, 2 = *somewhat safe*, and 3 = *very safe*. The violence question responses were collapsed into 1 = *very likely or likely*, 2 = *uncertain/don't know*, 3 = *unlikely*, and 4 = *very unlikely*.

Women's help seeking was measured with an inventory of formal and informal help seeking used by battered women (Gondolf & Fisher, 1988).

Previous contact with a women's shelter or calls to the police were used as separate variables because they represent more extreme help seeking. We also calculated the number of help-seeking strategies the woman pursued among the inventory items. Another variable we measured was the woman's seeking any form of legal assistance (e.g., protection order and pressing charges).

Program variables. The effects of program context were controlled by including the four sites—Pittsburgh, Denver, Houston, and Dallas—as a categorical variable. In all multivariate equations, the Pittsburgh site was the omitted reference group because it has the greatest control through 30-day mandatory court review, the shortest duration of treatment, and the least amount of services. Court-referred versus voluntary status was determined by the referral source indicated on the background questionnaire.

Risk Assessment Instruments

We were able to simulate three popular risk assessment instruments using our data: the K-SID, the SARA, and Campbell's DAS. The DAS was designed to assess the potential for lethality, whereas the K-SID and the SARA were designed to assess the risk for further violence and not necessarily lethality. Our "simulated" versions of the three risk assessment instruments are reasonable proxies for the actual instruments (i.e., they simulate actual use of the instruments), as they approximate the use of risk instruments in actual practice, where incomplete information is common. For example, a recent study using the SARA was able to obtain information from only about 30% of the victims (Williams & Houghton, in press).

The K-SID (Gelles & Tolman, 1998) is used widely in Connecticut for sentencing in domestic violence cases (Roehl & Guertin, 2000). It has three parts: the poverty chart, a severity and injury index, and an index of 10 risk factors. A study of the K-SID is currently under way by Eleanor Lyons in Connecticut, but prediction results are not available at this time. Our simulated version of the K-SID is excellent. The only item that is completely missing out of the 11 is "previous violation of a protection order." We have 7 of the 11 items available (age, batterer witnessed abuse, marital status, education, employment status, type of employment, and previous domestic violence arrest) and similar items for the remaining 3 items (poverty, drug use in past year, and child abuse or neglect allegations).

The DAS was developed by Jacqueline Campbell (1995) based on known risk factors for lethal violence. The scale consists of 15 yes-or-no questions; scoring is based on the total number of yes answers. Five studies have shown

very good reliability and moderate construct validity (Roehl & Guertin, 2000). A study by Goodman et al. (2000) showed that the DAS was modestly predictive of misdemeanor reassault in the short term. However, their original sample was quite small ($N = 92$), and their follow-up sample was even smaller ($n = 47$). Weisz et al. (2000), in their simulated DAS, found the DAS to be a modest predictor of reassault at a 4-month follow-up when used in conjunction with other significant predictors. Our simulated version of the DAS includes 13 of the 16 items. Eight are the same (use of or threaten with weapon, choking or attempted choking, forced sex, abuser uses drugs, abuser threatened to kill or survivor believes abuser to be capable of killing her, abuser drunk almost daily, abuser threatened or attempted suicide, and abuser violent toward children), and 5 are similar (increased frequency of physical violence in the past year, increased severity of physical violence in past year, abuser controls partner's daily activities, abuser violently and constantly jealous, and abuser violent outside the home). The 3 DAS items that were not available are presence of gun in the house, survivor beaten during pregnancy, and survivor threatened or attempted suicide.

The SARA, developed at the British Columbia Institute on Family Violence, consists of a clinical checklist of 20 risk factors, which are grouped into five areas: criminal history, psychosocial adjustment, spousal abuse history, current offense characteristics, and other (Kropp, Hart, Webster, & Eaves, 1999). Each of the 20 items is scored 0 to 2; risk is considered to increase with the number of items that are present (score of 2). The items are drawn from victim and offender interviews as well as criminal records. Although interrater reliability is high and internal consistency is moderate, evidence of predictive validity is modest. Only enhanced clinical judgments predict recidivism, and the prediction is modest (Kropp et al., 1999; Kropp & Hart, 2000). Preliminary results of a large-scale study in Colorado suggest that the SARA predicts reoffense better than chance (Dutton & Kropp, 2000). Our simulated SARA uses 16 of 20 items. Ten items are the same (recent relationship problems, recent employment problems, victim of and/or witness to family violence as a child or adolescent, recent substance abuse/dependence, recent suicidal or homicidal ideation/intent, past physical assault, past sexual assault or sexual jealousy, past use of weapons and/or credible threats of death, severe and/or sexual assault in most recent incident, and use of weapons and/or credible threats of death in most recent incident), and 6 items are similar (past assault of family members; past assault of strangers or acquaintances; recent psychotic and/or manic symptoms; personality disorder with anger, impulsivity, or behavioral instability; recent escalation in frequency or severity of violence; and extreme minimization or denial of spousal assault history). The 4 items that were unavailable were past viola-

tion of conditional release or community supervision, past violation of "no contact" orders, and violation of "no contact" order in most recent incident.

Analytic Procedures

To compare the prediction of women's perceptions compared to risk factors and instruments, we used multinomial logistic regression with our multiple outcome variable. The multivariate analyses were based on the 499 cases for which we have 15-month follow-up data. To model a multiple outcome variable with five categories, four equations that estimate log odds were computed. We used "repeat reassault" as the baseline category and determined how the predictor variables affected the log odds in each of the remaining four categories. The repeat reassault group was used as the baseline category because it is responsible for most of the severe and injurious assaults and therefore is of most concern. The model chi-square was used as a global test of the significance of the overall model.

To build the most efficient and powerful model for risk factors, we proceeded hierarchically, testing one risk factor at a time within blocks. We first entered referral status, age, and race as control variables that were kept in the models regardless of whether they were significant variables. We then entered social background variables one at a time and retained variables that had at least one significant coefficient at $p < .10$ (two-tailed) in any of the four equations. We also retained risk factors if the overall global test for the predictor was significant. In addition, we examined the classification tables available in SPSS and retained variables that improved overall classification or classification of repeat reassault.

The following blocks of variables were entered sequentially: socio-demographics, socioeconomic indicators, relationship variables, previous abuse and other related behaviors, personality and mental health variables, and program site. Once we developed a baseline risk factor model using these predictors, we then entered women's characteristics, such as help-seeking variables to see whether they contributed significantly to prediction. The last variables we entered were women's perceptions of risk for violence and feelings of safety to determine whether they improved prediction.

To compare women's perceptions to risk instruments, we ran bivariate multinomial logistic regression equations, in which we regressed the multiple outcomes on each of the risk instruments and women's perceptions variables separately. For each risk instrument (K-SID, SARA, and the DAS), we tried different coding strategies, such as total scores, number of factors present, and categorical risk categories (e.g., low, medium, and high risk) because these different coding strategies have been suggested in previous research.

To provide the fairest chance for the risk instruments to provide strong prediction, results are presented for the coding strategy that provided the greatest prediction. For further comparison, we also regressed the multiple outcomes on the DAS and SARA instruments in conjunction with each of the women's perceptions variables.

Only the strongest combined models are presented in the results, and comparisons across the various models were based primarily on the area under the curve (AUC) measures associated with receiver operating characteristics (ROC) curves (see Rice & Harris, 1995). Sensitivity (percentage of repeat reassault cases correctly classified), false positive rates (percentage of men wrongly classified as repeat reassault), and overall correct classification (percentage of accurate classifications across the five multiple outcomes) were also examined. Currently, the use of the ROC AUC statistic is the preferred summary measure in prediction research to evaluate how well a model or risk instrument predicts violence or dangerousness (Rice & Harris, 1995). Software and methods for calculating ROC curves (and AUC statistics) for multiple outcomes have not yet been developed. To calculate ROC curves, therefore, we used two strategies. One, we used the predicted probabilities for each individual's likelihood of being a repeat reassaulter (computed based on the multinomial logistic regression equations) as the "test" variable for the ROC curves. Two, we collapsed the multinomial outcome into two categories (repeat reassault versus the other four categories) and computed binary logistic regressions to predict repeat reassault and used the associated probabilities of being repeat reassaulters as the test variables for computing ROC curves. Because the results and conclusions were similar, we report the ROC AUC measures for the first method.

RESULTS

To show the percentage of women who felt at risk for future violence or who felt unsafe, Table 1 provides the frequency distributions for the two women's perceptions measures. The majority of women feel that violence is either unlikely (26%) or very unlikely (42%) and feel very safe (59%). The two risk variables are moderately correlated (Spearman's $r = .54$). They appear to be related but independent constructs.

Risk Factors for Repeat Reassault

A number of risk factors distinguished the repeat reassaulters (23% of the sample) from each of the other categories. Because the focus of this article is

TABLE 1: Frequency Distributions for Women's Perceptions of Risk at Program Intake (in percentages)

<i>Variable</i>	<i>Distribution</i>
Likelihood of violence in next 3 months (<i>n</i> = 542)	
Very likely or likely	12.4
Uncertain/don't know	19.6
Unlikely	25.6
Very unlikely	42.4
How safe woman feels at intake (<i>n</i> = 539)	
Uncertain, not safe, much danger	11.1
Somewhat safe	30.1
Very safe	58.8

on women's perceptions, however, we only briefly mention the significant risk factors, which are shown in Table 2. In line with previous research, important predictors included age, race, living together at intake, having children living with the batterer, the woman receiving injuries, previous psychological abuse, the batterer working as a laborer or service worker, having a non-domestic-violence arrest prior to intake, and use of a shelter or social services by the victim.

Women's Characteristics and Perceptions

We first assessed how much women's perceptions contribute to prediction of repeated reassault above and beyond men's characteristics and reports. Women's perceptions of risk substantially improved prediction in the multiple outcomes model over other risk factors. As Table 3 reveals, when only men's variables and reports are considered, the ROC AUC is .75 and the sensitivity rate of the multinomial outcomes model in predicting repeat reassault is 55%, with an overall classification rate of 38%. Inclusion of women's characteristics other than women's perceptions increased the ROC AUC to .79 and the sensitivity rate slightly to 58% (overall classification = 42%). Next, inclusion of women's perceptions of risk increased the ROC AUC to .83 and the sensitivity rate to 70% in the final model.

However, the relationship of women's perceptions to repeat reassault is not straightforward. As shown in Table 2, if the woman felt somewhat safe, the man was more likely to repeatedly reassault than if the woman felt very safe. However, if the woman did not know, felt not safe, or felt in much danger, the man was not significantly more likely to repeatedly reassault than if

TABLE 2: Unordered Multinomial Logistic Regression for Multiple Outcomes: Intake Model (n = 499)

Risk Marker	No Abuse vs. Repeat Reassault		Control/Verbal vs. Repeat Reassault		Threats vs. Repeat Reassault		One-Time vs. Repeat Reassault	
	b	OR	b	OR	b	OR	B	OR
Social background variables								
Age								
26-35 years	1.26***	3.54	1.02***	2.78	1.65***	5.23	0.75*	2.11
36-65 years	1.19***	3.30	1.17***	3.21	1.62***	5.05	0.88*	2.41
Race								
African American	0.49	1.63	0.86**	2.36	-0.16	0.85	1.01**	2.74
Latino	1.18**	3.27	1.19**	3.27	1.34**	3.82	0.73	2.07
Other	1.56**	4.74	0.76	2.13	0.58	1.78	0.31	1.36
Blue-collar occupation	0.52	1.68	0.61*	1.84	0.56	1.75	0.55	1.73
Relationship characteristics								
Living with partner	-0.87**	0.42	-0.65***	0.52	-1.25***	0.29	-0.54	0.58
Child(ren) living with batterer	0.92***	2.51	1.00***	2.71	0.95***	2.57	0.14	1.15
Previous behavior								
Verbal abuse in past 3 months	-1.66***	0.19	-1.21**	0.30	-1.47***	0.23	-0.65	0.52
Controlling behavior in past 3 months	-1.10***	0.33	-0.08	0.93	-0.22	0.80	-0.75*	0.47
Threats in past 3 months	-0.30	0.74	-0.99***	0.37	-0.34	0.71	-0.06	0.94
Non-domestic-violence arrest—ever	-0.14	0.87	-0.80**	0.45	-0.46	0.63	0.20	1.22
Woman injured	-0.96***	0.38	-0.63***	0.53	-0.74**	0.48	-0.56	0.57

TABLE 3: Correct Classification Using Risk Factors for Multiple Outcomes (*n* = 499)

<i>Variable</i>	<i>ROC AUC</i>	<i>% Sensitivity</i>	<i>% False Positives</i>	<i>% Overall Classification</i>
Men's characteristics and reports only	.75	55	20	38
Men's and women's characteristics and reports	.79	58	19	42
Men's and women's characteristics and reports, and women's perceptions of risk	.83	70	16	48

NOTE: ROC AUC = receiver operating characteristics area under the curve.

the woman felt very safe. In other words, when women feel greater risk, they are more likely to respond in ways to reduce that risk, such as leaving the batterer or engaging in safety planning. If they feel somewhat safe, they may feel uneasy or uncertain about the risk but not enough to take action.

In two of the four equations, if the woman was uncertain or did not know whether the man would use violence in the next 3 months, the man was more likely to be a repeat reassaulter over 15 months. Again, it is the uncertainty that may keep some women from taking actions to reduce the risk of reassault.

The second area we tested was the women's perceptions against simulated risk assessment instruments: the K-SID, SARA, and DAS. As Table 4 shows, the ability of the simulated K-SID total scores (and low, medium, and high risk categories, which are not shown) to accurately predict repeat reassault was extremely low (AUC = .57; sensitivity = 29%). The simulated SARA total scores predicted significantly better than the K-SID scores but still offered modest prediction of multiple outcomes (AUC = .64; sensitivity = 43%). Using a high-risk cutoff point for the SARA scores substantially reduced prediction from the SARA total scores, dropping the ROC AUC and the sensitivity rate (not shown).

The simulated risk assessment instrument that predicted best was the DAS, with a ROC AUC of .70. The simulated DAS scores based on women's reports also accurately classified a reasonably robust 66% of the repeat reassault, although the rate of false positives was 33%. The DAS more consistently distinguishes "repeat reassault" from each of the other four outcome categories, which helps explain why it has better overall prediction than the SARA. The DAS may distinguish "repeat reassault" from "one-time

TABLE 4: Correct Classification Using Simulated Risk Assessment Instruments and Women's Perceptions for Repeat Reassault (*n* = 499)

<i>Variable</i>	<i>ROC AUC</i>	<i>% Sensitivity</i>	<i>% False Positives</i>	<i>% Overall</i>
K-SID total score	.57	11	5	27
SARA total score	.64	43	27	32
DAS, women's reports	.70	66	33	33
Women perceive violence as likely	.64	52	26	32
Women's perceptions of safety	.63	63	40	31
DAS and violence likely	.73	63	27	35
DAS and safety	.72	64	28	37
SARA and violence likely	.69	54	26	34
SARA and safety	.67	55	31	34

NOTE: ROC AUC = receiver operating characteristics area under the curve.

reassault" better than the SARA because it was developed to predict lethal violence, whereas the SARA was designed to predict reassault in general.

The women's perceptions of risk were then entered individually in multinomial logistic regression equations with no controls (see Table 4). Interestingly, women's perceptions of safety and how likely the man is to hit (assessed at intake) were also modest predictors of multiple outcomes by themselves (AUC = .64 for violence likely and AUC = .63 for perceptions of safety). Women's perceptions predicted better than the K-SID (AUC = .57), similar to the SARA (AUC = .64), and not as well as the DAS (AUC = .70). Feelings of safety at intake accurately classified 64% of "repeat reassault," which was almost as high a rate as the DAS, although the rate of false positives was nearly 40%. Women's perceptions of how likely the man was to use violence had a reasonably high sensitivity rate of 55%, with a relatively low rate of false positives (26%). The combination of the DAS and women's perceptions that future violence was likely was the best model for the instruments (AUC = .73; sensitivity = 64%) but was not as predictive as the equations with individual risk factors when women's characteristics and women's perceptions were included (AUC = .83; sensitivity = 70%).

For comparison, we conducted the same analyses with binary logistic regression, using "any reassault" over 15 months as the outcome (results not shown). These analyses allowed comparisons to previous risk assessment research, which typically uses a dichotomous outcome. The patterns with regard to which risk assessment instruments and items provide the best prediction were virtually identical to the multinomial results.

DISCUSSION

Summary

Our analysis of women's perceptions of risk compared to risk factors and risk instruments confirmed the contribution of women's perceptions to the prediction of especially repeat reassault. Women's perceptions of risk substantially improved prediction of repeat reassault (AUC = .83 vs. .79; sensitivity improved 12 percentage points) over the prediction with men's and women's characteristics alone. The relationship of women's perceptions to repeat reassault, however, was not straightforward. Generally, if women felt somewhat safe, they had a higher risk of repeat reassault than if they felt very safe. In addition, if women were uncertain whether the man would use violence, the man was more likely to repeatedly reassault.

The prediction with the K-SID had poor predictive ability, and the simulated SARA had better, but still modest, prediction. The DAS predicted repeat reassault almost as well as the risk factor model (inclusive of women's perceptions) but had a higher rate of false positives and a lower ROC AUC (.70 vs. .83). Women's perceptions of risk by themselves were much better predictors than the simulated K-SID, similar to the SARA, and not quite as accurate as the DAS. Women's perceptions of risk as predictors did produce relatively high rates of false positives, which lowered their ROC AUCs. The best models were the risk factor models, particularly the one that included women's perceptions, followed by the model that included the DAS and women's perceptions of risk.

In sum, women's perceptions of risk are a reasonably accurate predictor of repeated reassault by themselves and improve the prediction of risk factors and instruments. The fact that women's perceptions of risk appear to enhance risk assessment efforts suggest that women may be able to assess risk in broader and more idiosyncratic ways that go beyond a combination of risk factors.

Limitations

Although our study addressed many of the methodological limitations facing prediction research, there are several limitations to the current study. First, reliability and validity of measurement, especially the outcome measure, are always concerns in domestic violence research. Despite measurement limitations, the data set we used is one of the best databases available with regard to measurement, owing to the extent and variety of variables and the short duration between interviews. The data are much of what is practi-

cally obtainable by clinicians and what is most relevant to them. More sophisticated measurement would mean developing instrumentation and testing that is beyond the current state in the criminal justice system.

The quality and comprehensive nature of the data set also offset, to some extent, the limitations that the risk instruments (DAS, SARA, and K-SID) were simulated using the data rather than administered in actual clinical practice. Our data set probably has less missing data than most "real world" administrations of risk instruments. We contend that our "simulated" versions of the three risk assessment instruments are reasonable proxies for the actual instruments in that they simulate actual clinical usage.

The prediction models may also be limited by the clinical nature of the sample. They apply to batterer program outcome, and not the potentiality of reassault. The men's potential for reassault is mediated by the batterer programs and additional services and interventions in the four sites. Some men for whom the risk factors would have predicted repeat reassault were contained by the system. We would argue, however, that prediction of program outcomes, rather than potentiality, is more relevant to policy development. We want to know which men batterer counseling is not containing and who, therefore, need additional or different interventions.

Implications

First, the simulated risk assessment instruments appeared to offer only modest prediction in our study and should be used with caution by batterer programs and the criminal justice system, as previous research has already recommended (see Roehl & Guertin, 2000). Results are improved somewhat with additional items and women's perceptions, reinforcing the importance of using instrument results in combination with a variety of other sources of information. Second, the predictive power of women's perceptions suggests the importance of obtaining and heeding women's appraisal of their situation, as advocates have long argued, and including them in risk instruments. Batterer program staff and the courts may have to work more closely with women's program staff to obtain such information and incorporate it into their assessments.

Third, regarding feelings of safety, the women who are at greatest risk may be those who feel somewhat safe. This may be because they have some uncertainty or uneasiness but not enough to take proactive action to reduce risk. This finding supports De Becker's (1997) argument that women tend not to trust their "intuition." If the women do not assert that they feel very safe, it would be reasonable to inform them to trust their instincts if they feel even slight uncertainty. They should be counseled to pay close attention to

any warning signs of impending assault. In addition, a subset of women may put themselves at greater risk because they perceive violence as very unlikely and therefore do not take necessary precautions. This finding may also reflect the very dynamic nature of some of these relationships.

Fourth, further research needs to explore in more depth the precise nature of the relationship between women's perceptions and reassault as well as the impact of women's perceptions on their safety-related actions. Models should also be developed to see how well women's perceptions predict reassault in the short term (e.g., over 3 months) as opposed to the models that we developed that predicted reassault over 15 months. Future research using an independent sample is also needed to replicate the results of the current study. A cross-validation procedure was not feasible for the current study because multinomial logistic regression requires a larger sample size for cross-validation than was available.

Our current study, overall, confirms the contribution of women's perceptions to the prediction of reassault. In the process, it reinforces the importance of these perceptions in assessing risk and prescribing intervention in domestic violence cases.

REFERENCES

- Blackburn, R. (1998). Psychopathy and the contribution of personality to violence. In T. Millon, E. Simonsen, M. Birket-Smith, & R. Davis (Eds.), *Psychopathy: Antisocial, criminal, and violent behavior* (pp. 50-68). New York: Guilford.
- Campbell, J. (1986). Assessment of risk of homicide for battered women. *Advances in Nursing Science*, 8, 36-51.
- Campbell, J. (1995). Prediction of homicide of and by battered women. In J. Campbell (Ed.), *Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers* (pp. 96-113). Thousand Oaks, CA: Sage.
- De Becker, G. (1997). *The gift of fear: Survival signals that protect us from violence*. New York: Dell.
- DeMaris, A., & Jackson, J. K. (1987). Batterers' reports of recidivism after counseling. *Social Casework*, 68, 458-465.
- Dutton, D. (1986). The outcome of court-mandated treatment for wife assault: A quasi-experimental evaluation. *Violence and Victims*, 1, 163-175.
- Dutton, D., Bodnarchuk, M., Kropp, R., Hart, S., & Ogloff, J. (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., & Kropp, P. (2000). A review of domestic violence risk instruments. *Trauma, Violence & Abuse*, 1, 171-181.
- Gelles, R., & Tolman, R. (1998). *The Kingston Screening Instrument for Domestic Violence (K-SID)*. Unpublished risk instrument, University of Rhode Island, Providence.
- Gondolf, E. (1997). Patterns of reassault in batterer programs. *Violence and Victims*, 12, 373-387.

- Gondolf, E. (1999). Characteristics of court-mandated batterers in four cities. *Violence Against Women, 5*, 1277-1293.
- Gondolf, E. (2003). *Batterer types based on the MCMI: A less than promising picture*. Manuscript submitted for publication.
- Gondolf, E., & Fisher, E. (1988). *Battered women as survivors: An alternative to treating learned helplessness*. Lexington, MA: Lexington Books.
- Gondolf, E., Heckert, A., & Kimmel, C. (2002). Non-physical abuse among batterer program participants. *Journal of Family Violence, 17*, 293-314.
- Gondolf, E., & White, R. (2001). Batterer program participants who repeatedly reassault: Psychopathic tendencies and other disorders. *Journal of Interpersonal Violence, 16*, 361-380.
- Goodman, L., Dutton, M., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers. *Journal of Interpersonal Violence, 15*, 63-74.
- Hamberger, L. K., & Hastings, J. E. (1988). Skills training for treatment of spouse abusers: An outcome study. *Journal of Family Violence, 3*, 121-130.
- Hamberger, L. K., & Hastings, J. E. (1990). Recidivism following spouse abatement counseling: Treatment program implications. *Violence and Victims, 5*, 157-170.
- Hamberger, L. K., Lohr, J., Bonge, D., & Tolin, D. (1996). A large sample empirical typology of male spouse abusers and its relationship to dimensions of abuse. *Violence and Victims, 11*, 277-292.
- Hanson, R. K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers* (User Report 2000-06). Ottawa: Department of the Solicitor General of Canada.
- Heckert, A., & Gondolf, E. (2000a). Assessing assault self-reports by batterer program participants and their partners. *Journal of Family Violence, 15*, 181-197.
- Heckert, A., & Gondolf, E. (2000b). Predictors of underreporting of male violence by batterer program participants and their partners. *Journal of Family Violence, 15*, 423-443.
- Heckert, A., & Gondolf, E. (2002). *Predicting levels of abuse and reassault among batterer program participants* (Final report). Washington, DC: National Institute of Justice.
- Heckert, A., & Gondolf, E. (2003). *Do multiple outcomes and conditional factors improve prediction of batterer reassault?* Manuscript submitted for publication.
- Hilton, N. Z., Harris, G., & Rice, M. (2001). Predicting violence by serious wife assaulters. *Journal of Interpersonal Violence, 16*, 408-423.
- Holtzworth-Munroe, A., & Stuart, G. L. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin, 116*, 476-497.
- Jones, A. (2002). *Sample attrition in panel data from an evaluation of four batterer intervention programs*. Manuscript submitted for publication.
- Kropp, P., & Hart, S. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior, 24*, 101-118.
- Kropp, P., Hart, S., Webster, C., & Eaves, D. (1999). *Spousal Assault Risk Assessment guide: User's manual*. Vancouver, Canada: Multi-Health Systems Inc.
- Limandri, B., & Sheridan, D. (1995). Prediction of intentional interpersonal violence. In J. Campbell (Ed.), *Assessing the risk of dangerousness: Potential for further violence of sexual offenders, batterers, and child abusers* (pp. 1-19). Thousand Oaks, CA: Sage.
- Millon, T. (1994). *Millon Clinical Multiaxial Inventory-III manual*. Minneapolis, MN: National Computer Systems.
- Millon, T., & Davis, R. (1998). Ten subtypes of psychopathy. In T. Millon, E. Simonsen, M. Birket-Smith, & R. Davis (Eds.), *Psychopathy: Antisocial, criminal, and violent behavior* (pp. 161-187). New York: Guilford.
- Mulvey, E., & Lidz, C. (1993). Measuring patient violence in dangerousness research. *Law and Human Behavior, 17*, 277-278.

- Rice, M., & Harris, G. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.
- Roehl, J., & Guertin, K. (2000). Intimate partner violence: The current use of risk assessments in sentencing offenders. *Justice System Journal, 21*, 171-197.
- Saunders, D. (1995). Prediction of wife assault. In J. Campbell (Ed.), *Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers* (pp. 68-95). Thousand Oaks, CA: Sage.
- Saunders, D. (1996). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and Victims, 11*, 393-414.
- Selzer, M. (1971). The Michigan Alcoholism Screening Test (MAST): The quest for a new diagnostic instrument. *American Journal of Psychiatry, 3*, 176-181.
- Shepard, M., & Campbell, J. (1992). The Abusive Behavior Inventory: A measure of psychological and physical abuse. *Journal of Interpersonal Violence, 7*, 291-305.
- Spradley, J. (1979). *The ethnographic interview*. New York: Holt, Rinehart & Winston.
- Steadman, H., Monahan, J., Appelbaum, P., Grisso, T., Mulvey, E., Roth, L., et al. (1994). Designing a new generation of risk assessment research. In J. Monahan & H. Steadman (Eds.), *Violence and mental disorder: Developments in risk assessment* (pp. 297-318). Chicago: University of Chicago Press.
- Straus, M. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41*, 75-88.
- Tolman, R. (1989). The development of a measure of psychological maltreatment of women by their male partners. *Violence and Victims, 4*, 159-177.
- Tolman, R., & Bennett, L. (1990). A review of quantitative research on men who batter. *Journal of Interpersonal Violence, 5*, 87-118.
- Weisz, A., Tolman, R., & Saunders, D. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence, 15*, 75-90.
- Williams, K. R., & Houghton, A. B. (in press). Assessing the risk of domestic violence re-offending: A validation study. *Law and Human Behavior*.

D. Alex Heckert is a professor of sociology at Indiana University of Pennsylvania. His primary areas of teaching and research interest are family sociology, deviance, and medical sociology. His collaborative research in the area of domestic violence interfaces these areas of interest. Recent research has attempted to improve prediction of nonphysical abuse and physical reassault among batterer program participants. He has previously published in journals such as Deviant Behavior, Journal of Research in Crime and Delinquency, Journal of Marriage and Family, Demography, Journal of Family Issues, Rural Sociology, Family Relations, Journal of Family Violence, Women and Criminal Justice, Research on Aging, and Journal of Elder Abuse and Neglect.

Edward W. Gondolf, Ed.D., MPH, is research director for the Mid-Atlantic Addiction Training Institute (MAATI) and professor of sociology at Indiana University of Pennsylvania. He conducts grant-funded research on the response of the courts, mental health practitioners, alcohol treatment clinicians, and batterer treatment programs to domestic violence. He is the author of several articles and books on these topics, including Assessing Women Battering in Mental Health Settings (Sage, 1998) and Batterer Intervention Systems: Issues, Outcomes, and Recommendations (Sage, 2002).