Law and Human Behavior
2012, Vol. 36, No. 2, 120–129

0147-7307/11/$12.00 DOI: 10.1037/h0093977

Family Violence Risk Assessment: A Predictive Cross-Validation Study of the Domestic Violence Screening Instrument-Revised (DVSI-R)

Kirk R. Williams
University of California

This research was a cross-validation study of the Domestic Violence Screening Instrument-Revised (DVSI-R), using a diverse, statewide sample of 3,569 family violence perpetrators in Connecticut, assessed in February and March of 2007. It analyzed re-arrest data collected during an 18-month period post assessment. Three issues were central, which have been ignored in previous research on family violence risk assessment: (1) analyzing five refined measures of behavioral recidivism, (2) determining whether perpetrator characteristics and types of family and household relationships (beyond just heterosexual intimate partners) moderate the empirical relations between the DVSI-R and the behavioral recidivism measures, and (3) determining whether structured clinical judgment about the imminent risk of future violence to the victim or to others corresponds with recidivism predicted by the DVSI-R total numeric risk scores. The empirical findings showed that the DVSI-R had significant predictive accuracy across all five measures of recidivism. With one exception, these relations did not vary by gender, age, or ethnicity; and again with one exception, no significant evidence was found that types of family or household relationships moderated those empirical relations. In short, the evidence suggested that the DVSI-R was a robust risk assessment instrument, having applicability across different types of perpetrators and different types of family and household relationships. Finally, the empirical findings showed that structured clinical judgment about imminent risk-to-victim and risk-to-others corresponded with the prediction of recidivism by the DVSI-R total numeric risk scores, but the effects of those scores were significantly stronger than the perceived risk-to-victim or the perceived risk-to-others.

Keywords: risk assessment, family violence, domestic violence screening

Legislative changes governing legal responses to family violence have swept the U.S. during the last 30 years. Virtually every state in the country has expanded police authority to make warrantless arrests with probable cause that a family violence crime has been committed, and 23 states now have mandatory arrest laws, with six states having preferred arrest laws for such crimes (Hirschel, Buzawa, Pattavina, Faggiani, & Reuland, 2007). Besides the expansion of police authority, the scope of relationships involved in family violence crimes has been substantially broadened beyond married couples and their children to a host of other relationships classified as family or household members (e.g., unmarried couples residing together, persons involved in dating relationships, or roommates). The scope of the behavior covered under family violence laws has also been extensively broadened (e.g., reckless endangerment, kidnapping, breach of peace, various degrees of assault, or stalking). One upshot of these legislative changes has been a burgeoning number of arrests for family violence crimes (e.g., Hirschel et al., 2007; National Research Council, 2004; Simpson, Bouffard, Garner, & Hickman, 2006). The challenge facing the criminal justice system is how to assess and manage an increasingly heavy caseload.

This challenge has been met at least in part by a parallel development, specifically, the construction of risk and needs assessment instruments to guide the management of criminal offenders, particularly their supervision and treatment (e.g., Andrews & Bonta, 2006; Brennan, Dieterich, & Ehret, 2009). Some instruments have been generic, focusing on overall criminal risk and needs of youth or adult offenders (e.g., COMPAS, Brennan et al., 2009; LSI-R, Andrews & Bonta, 1995; LS/CMI, Andrews, Bonta, & Wormith, 2004). Other instruments have been more specific, focusing on a particular type of behavior, such as violence (e.g., VRAG, Quinsey, Harris, Rice, & Cormier, 2006) or a subtype of that behavior, such as intimate partner violence (for reviews, see Dutton & Kropp, 2000; Hilton & Harris, 2005; Hilton, Harris, & Rice, 2010). The development of family violence risk assessment instruments has been accompanied by empirical investigations of their predictive validity (Goodman, Dutton, & Bennett, 2000; Heckert & Gondolf, 2004; Williams & Grant, 2006; Williams & Houghton, 2004). The research aim of these investigations has been to determine whether *assessed* risk predicts *actual* risk, as measured by behavioral recidivism (i.e., repeated criminal or violent behavior post assessment).

This article was published Online First March 15, 2011.

Kirk R. Williams, Presley Center for Crime and Justice Studies, University of California, Riverside, Riverside, CA.

This research was conducted with assistance from the Judicial Branch, Court Support Services Division (CSSD) of the State of Connecticut. Special appreciation is expressed to William Carbone, Executive Director of CSSD. Appreciation is also expressed (in alphabetical order) to Kathy Ceruti, Joe DiTunno, Steve Grant, Brian Hill, Deb Kulak, and Brian Sperry.

Correspondence concerning this article should be addressed to Kirk R. Williams, Presley Center for Crime and Justice Studies, University of California, Riverside, Riverside, CA 92521, USA. E-mail: Kirk.Williams@ucr.edu

This study empirically evaluated the predictive validity of a specific family violence risk assessment instrument—the DVSI-R (Williams, 2008; Williams & Grant, 2006). This instrument is a modified version of the Domestic Violence Screening Instrument (DVSI) constructed and validated in Colorado (Williams & Houghton, 2004, and see Williams, 2008 for a discussion of the developmental history of the DVSI and DVSI-R). Other researchers have also analyzed the predictive validity of the original DVSI, reporting associations with recidivism in the range of other risk assessment instruments (e.g., Hilton, Carter, Harris, & Sharpe, 2008).

However, only a single predictive validity study of the DVSI-R has been conducted (Williams & Grant, 2006). It involved an analysis of 14,970 assessments conducted during September of 2004 and May of 2005 in Connecticut (Williams & Grant, 2006). The empirical results showed the DVSI-R predicted recidivism irrespective of the gender, age, and ethnicity of the perpetrator, or the nature of the intimate relationships under study (i.e., intimate partners, parents and children, or other family and household relationships). However, that study had two major limitations: (1) it was based on the developmental sample used to construct the instrument, and (2) the recidivism measure used was a proxy—number of times perpetrators were returned to authorities for risk assessments during the 8-month period (instead of re-arrests for repeated offending).

The Present Study

The earlier limitations of the initial research on the predictive validity of the DVSI-R were addressed in this study by using an independent, cross-validation sample of 3,569 family violence perpetrators in the state of Connecticut, assessed in February and March of 2007. Further, unlike previous research on family violence risk assessment, five refined measures of recidivism were analyzed, as described shortly. Re-arrest data were collected over an 18-month period following the dates the 3,569 cases were assessed during February and March of 2007. The robustness of the DVSI-R was also empirically evaluated by determining whether empirical relations between the total numeric risk scores and the five behavioral recidivism measures were moderated by perpetrator characteristics (age, gender, ethnicity, etc.) or different types of family and household relationships (intimate partners, parents and their children, and other types of family and household members). Estimation of such moderating effects has not been addressed in previous predictive validation studies of family violence risk assessment instruments. In addition, as described below, the DVSI-R includes a mechanism for clinically overriding the total numeric risk score. Evaluators responsible for the risk assessments are asked to judge the perceived imminent (within the next 6 months) risk of future violence to the victim or other persons known to the victim and/or perpetrator. Such clinical judgment is "structured" because it is informed by the numeric scoring of the DVSI-R, but as shown shortly, although the perceived imminent risk ratings were correlated with the total numeric risk scores, the correlations were clearly not perfect. Accordingly, this study empirically examined whether the prediction of perceived imminent risk-to-victim or perceived imminent risk-to-others corresponded with the prediction of risk by the DVSI-R total numeric risk scores. This issue also has not been addressed in any previous research on family violence risk assessment.

Risk Assessment Instruments and Predictive Validity

Several family violence risk assessment instruments have been developed and evaluated empirically, with results being supportive, although estimate effects of risk scores on recidivism (usually re-arrest for family violence in general) have varied in magnitude across studies (for reviews, see Dutton, 2006; Hilton et al., 2008). Risk assessment instruments have also varied in the purpose of their use. For example, one instrument was designed to facilitate relatively quick frontline assessments by police concerning arrest decisions (ODARA). This instrument was based on the use of information readily available at the time the assessments were conducted by police officers. Other instruments were designed for more in-depth follow-up assôessments (e.g., SARA, Kropp & Hart, 2000; DVRAG, Hilton et al., 2008).

A risk assessment system has been proposed, which combines a frontline assessment using the ODARA with an in-depth assessment using the DVRAG (Hilton et al., 2010). These more in-depth follow-up assessments typically require more time to administer and more expertise on the part of the evaluator. For example, the DVRAG combines the ODARA with the Psychopathy Check List-Revised (PCL-R), developed by Hare (1991). The administration of this instrument requires a more advanced level of clinical training than is typical of most criminal justice practitioners. Some components of the SARA also require similar clinical expertise (e.g., in diagnosing recent psychotic or manic disorders, personality disorder with anger, impulsivity, and behavioral instability). Another instrument was designed as a lethality assessment instrument administered to victims, not perpetrators of intimate partner violence (Dangerousness Guide, Campbell, 2007). These two distinctions (lethality and a victim administration) differentiate it from all other risk assessment instruments, although the DVRAG has been shown to be significantly associated with the severity of victim injury (Hilton et al., 2008).

Like the ODARA, the DVSI-R is a frontline assessment instrument that can be administered relatively quickly. Unlike the ODARA, the DVSI-R has been administered by Family Relations Counselors in Connecticut and informs recommendations they make to the court at arraignment, not arrest decisions by police. These counselors are trained evaluators (84% have Master's Degrees or higher) within a special division (Family Services) of the Court Support Services Division of the Judicial Branch of government in Connecticut. Besides assessing and administratively monitoring all family violence cases arrested in this state, they conduct counseling and mediation services in civil cases involving other family matters. Persons arrested on family violence charges in Connecticut are required to be arraigned the next day court is in session. During the approximately 24-h period between arrest and the initial court appearance, Family Relations Counselors conduct a pre-arraignment assessment of all the cases entering the system, typically under chaotic circumstances (e.g., sometimes in jail settings, sometimes in cramped court offices, with people coming and going, sometimes in a copy room where others are busy making copies of records, etc.). Like the legislative changes mentioned at the outset, the statutory definition of "family and household members" in Connecticut is quite broad, covering a diverse set of

relationships. It is presented here to show the expansiveness of family and household relationships covered by the statute and to justify the measurement of the relationship types, discussed below:

> Family/household members include (a) spouses, former spouses, (b) parents and their children, (c) persons eighteen years of age or older related by blood or marriage, (d) persons sixteen years of age or older other than those persons in subparagraph (c) presently residing together or who have resided together, (e) persons who have a child in common regardless of whether they are or have been married or have lived together at any time, and (f) persons in, or have recently been in, a dating relationship (CGS 46b-38a(2).

Hence, the risk assessment instrument in this state (or other states now having broad definitions of who qualifies for family violence crimes) must be applicable across violence within these diverse relationships. This issue has not been addressed in any previous research focusing exclusively on heterosexual intimate partner violence, with men as perpetrators and women as victims.

The assessments conducted with the DVSI-R in Connecticut guide recommendations made by Family Relations Counselors to the court on placement of the case and protective orders. Possible recommendations about placement include referral to family services for pre-trial supervision, referral to the State's Attorney for prosecution, *nolle* (no prosecution, but the case stays open for 13 months and if violations occur, the case can be reopened, with prosecution of the original and new offenses being possible), continued monitoring to ensure compliance with court orders, or dismissal. Recommendations for protective orders include no protective order, a partial or limited protective order, a residential stay away, or a full no contact protective order.

The preceding discussion suggests that several considerations should be taken into account as criminal justice practitioners and service providers evaluate alternative family violence risk assessment instruments. These considerations include the conditions under which assessments are conducted (chaotic and stressful vs. more regulated and relaxed), the time constraints of those assessments, whether perpetrators and/or victims are to be assessed, the expertise required of those administering the assessments, and the intended use of the risk assessment outcomes (arrest decisions, placement recommendations to the court, recommendations concerning protective orders, level of supervision for cases on probation, safety planning for victims at risk of lethality, etc.). Risk assessment instruments vary along these dimensions. Despite this variability, predictive validity is required of all risk assessment instruments, with confidence in this matter strengthened through accumulated evidence across multiple empirical studies. This point underscores the objectives of this predictive validation study of the DVSI-R.

Method

Sample Selection

Following the normal operating procedures of processing family violence cases in Connecticut, a cross-validation sample was generated that included 3,569 perpetrators assessed between February and March of 2007. All the cases assessed during this 2-month period formed the basis of the sample. An 18-month follow-up period was used to collect data on family violence recidivism, with that period beginning from the date the risk assessment was conducted and ending at 18 months after that date. The sample covered all of the 23 judicial geographic areas in Connecticut, meaning it represented the full population of defendants 16 years of age or older for the entire state during the assessment and follow-up period.

Risk Assessment Instrument: DVSI-R

The DVSI-R includes 11 items, seven of which primarily address the behavioral history of perpetrators (prior non-family assaults, arrest, or criminal conviction; prior family violence assaults, threats, or arrests; prior family violence intervention or treatment; prior violation of orders of protection or court supervision; prior or current verbal or emotional abuse; the frequency of family violence in the previous 6 months; and escalation of violence in the previous 6 months). The other four items include substance abuse, weapons use, children present during violent incidents, and employment status. Item scores range from 0 to 2 or 0 to 3, depending on the item, estimating the intensity of the risk factor, not just whether it is present for the individual being evaluated. For example, one item on the instrument is "evidence of non-family assaults, arrests, or criminal convictions," and the scoring categories are "0 = no evidence, 1 = one or two incidents, 2 = three or more incidents." An example of an item with a 0–3 scoring category is "history of violation of orders of protection or court supervision," and the scoring categories are "0 = no evidence, 1 = prior incident, 2 = current incident, 3 = prior and current incidents." The possible range of the total numeric risk scores generated by the administration of the DVSI-R is 0–28.

Family Relations Counselors administer intake interviews shortly after an offender has been arrested. They code the items during those interviews in which the assessments are conducted, and the total numeric risk score is automatically generated by a computerized system. The assessments draw from five sources of information: defendant interviews, a review of police reports, criminal history reviews, protective and restraining order registry reviews, and victim interviews usually conducted by victim advocates, with information passed on to Family Relations Counselors if the victim gives consent. In short, although performed quickly and typically under chaotic conditions, the DVSI-R is informed by a relative large array of information sources, unlike the original version of this instrument developed in Colorado (Williams & Houghton, 2004). It relied solely on a criminal record review by probation officers and was used to guide their recommendations to judges about the placement of male perpetrators of wife assault (private vs. state supervision).

This researcher trained the Family Relations Counselors on risk assessment in general and the DVSI-R in particular, with booster trainings occurring periodically by either the researcher or supervisors within the Family Relations Counselors offices. All the 23 geographic areas in Connecticut have such offices. A detailed discussion is provided elsewhere (to avoid unnecessary duplication in this article) of the history of family violence legislation and accompanying changes in policy and practice concerning the processing of family violence cases in Connecticut, the various activities performed during the stages of developing the DVSI-R in this state, and other training as well as psychometric properties (Wil-

liams, 2008; Williams & Grant, 2006; Williams & Houghton, 2004).

The DVSI-R also includes two perceived imminent risk ratings for evaluators to provide their professional assessment of cases after the 11 items are coded and a total numeric risk score is calculated. One addresses the perceived imminent risk of violence to the victim of the incident for which an arrest was made. The other addresses the perceived imminent risk to another person known to the victim and/or perpetrator. The scoring categories for these ratings are low (0), moderate (1), or high risk (2). Both of these ratings were included in the analyses to determine their estimated effects on behavioral recidivism, compared to the total numeric risk score generated by the 11 items included in the DVSI-R.

The rationale for including the perceived imminent risk ratings in this instrument was to provide a mechanism for evaluators to override the total numeric risk score if they identified information beyond the 11 DVSI-R items that would raise questions about its accuracy. To illustrate, during the initial developmental period, a Family Relations Counselor indicated that he had assessed an individual with the instrument, with that person receiving a total risk score of "1" (a very low score). He had one prior arrest and conviction for a family violence incident that occurred many years before the current offense. However, the incident involved a seriously violent episode in which he almost killed his former girlfriend. Given this information, the counselor did not consider the score as accurately representing the risk this individual posed to the current victim. The perceived imminent risk ratings, therefore, allowed an override of the total numeric risk score to guide his recommendations to the court. Regardless, the total numeric risk scores were highly correlated with the perceived imminent risk to victim ($r = .73$) and to others ($r = .61$), suggesting considerable agreement between the numeric assessment of recidivistic risk and perceptions of that risk. Nonetheless, those correlations were clearly not perfect, suggesting some independent input based on the counselors' structured professional judgment.

Measures of Family Violence Recidivism

Multiple measures of behavioral recidivism were used in the analysis because finding similar results across different measures enhances confidence in the predictive validity of the DVSI-R. Such results also document the robust nature of the instrument; that is, the utility of the DVSI-R to predict diverse outcomes. Re-arrest data were used to measure recidivism. The primary limitation of these data is well known. Not all reoccurring incidents of family violence are detected and reported to the police, and despite Connecticut being one of the 23 states having a mandatory arrest law, not all reported incidents always result in an arrest on an official offense charge. However, these data are commonly used in recidivism studies, with alternatives such as perpetrator or victim self-reports being too costly, ethically controversial, and having their own sources of error (for a more general discussion see Hilton et al., 2010; Williams, Tuthill, & Lio, 2008).

Five measures of recidivism during the 18-month follow-up period were used: (1) re-arrests for new family violence offenses only, (2) re-arrests for violations of protective or restraining orders only, (3) re-arrests for new family violence offenses *and* violations of protective or restraining orders only, (4) re-arrests for any family violence offense or violations of court orders (i.e., measures 1–3 combined as an overall family violence recidivism measure), and (5) total re-arrests (i.e., all offenses and violations combined whether family related or not).

The rationale for using the first four measures should be self-evident—empirically determining whether *assessed* recidivistic risk of family violence was associated with *actual* recidivistic risk of this behavior. The "only" term used with the four family violence recidivism measures denotes that each of these categories included "refined" types, meaning offenders re-arrested for some type of family violence recidivism (capture in the overall family violence recidivism measure, i.e., number 4 above) were sorted into the specific and mutually exclusive categories enumerated above. Dummy variables were calculated in which those falling into one of these categories were scored "1," and those not re-arrested for anything were scored "0." The overall measure of family violence recidivism was also a dummy variable, where scoring on any of the three types of family violence recidivism was included in the "1" category, and not being re-arrested for anything was scored "0."

The rationale for constructing these more refined measures is twofold. First, a major contribution of this study, as noted above, is that it estimated the accuracy of the DVSI-R in predicting violations of court orders in addition to family violence re-offending. Such analyses have not been performed in previous family violence risk assessment studies, and finding predictive effects could guide court personnel in determining both the level of such orders (e.g., a limited vs. a no contact protective order) as well as the level of monitoring needed while protective or restraining orders are in place.

Second, recent research has shown that using "ambiguous" cases in the measurement of family violence recidivism will result in an underestimation of the recidivism rate and the predictive accuracy of risk assessment instruments, yielding "risk assessment systems that mischaracterize risk" (Hilton & Harris, 2009, p. 328). For example, many participants in the present sample were not re-arrested for anything during the 18-month follow-up period (47.86%), and thus, were legitimately scored "0." However, a non-trivial percentage were re-arrested on other offenses during this period, including violent offenses, but were not classified as a family violence crimes (13.62%). Such cases can introduce error in estimation.

Total re-arrests for all types of re-offending (re-arrested for any family or non-family offense = 1 and not re-arrested for anything = 0) were used to determine whether the predictive accuracy of the DVSI-R was maintained when these more ambiguous cases were combined with those clearly arrested for a new family violence offense or violations of protective or restraining orders. Besides the inclusion of the family violence re-arrests, an association between the risk scores and re-arrests for any offenses would be expected because the first item on the DVSI-R bears directly on evidence of previous non-family offenses, and two other items could also be associated with non-family offending—violations of court orders or court supervision, some of which may have nothing to do with family violence (e.g., terms of probation or parole, bail, etc.) and employment status. Hence, scoring on these items should be associated with other types of re-offending during the follow-up period. However, to the extent these cases are not confounded with

family violence re-offending, that association should be attenuated, compared to the estimated effect of the total numeric risk scores on the four specific measures of family violence recidivism.

Family and Household Relationship Types

As noted above, previous studies of family violence risk assessment have focused exclusively on heterosexual intimate partner violence. This study broadened the focus by empirically examining whether the predictive accuracy of the DVSI-R differed by family and household relationship types. Three types consistent with the Connecticut statutory definition of family and household members (see above) were identified. Intimate partner included current or former spouses; unmarried couples residing together currently or previously, those currently or previously involved in dating relationships, or those having a child in common. Parent–child relationships included incidents in which the violence was directed from the parent to the child or the child to the parent regardless of the age of the perpetrators and victims. Hence, this category could include child abuse, elder abuse, or violence involving parents and their children when they are adults. A final category of "other types" was used that included a wide variety of family or household members, such as aunts, uncles, cousins, siblings, or roommates, as long as they were 16 years of age or older. Dichotomous dummy variables were constructed for each type, with a score of "1" assigned to cases that were included in a particular type (e.g., intimate partner), and a score of "0" assigned to cases that were not included in that particular type.

Characteristics of Perpetrators

Besides determining whether estimated effects of the DVSI-R total numeric risk scores on the recidivism measures differed by family and household relationship types, differences by characteristics of perpetrators were also empirically examined, another issue not addressed in previous research. Three characteristics were included in the analyses: gender (0 = females, 1 = males), age split at the mean to form two categories (0 = 18–35 and 1 = 36–84), and ethnicity. The latter characteristic was dichotomized, with Non-Latino Whites scored as "1," and all other ethnic minorities scored as "0." This dummy variable, therefore, compares an ethnic majority to ethnic minorities within this Connecticut-based sample.

Analytic Plan

Receiver operating characteristic (ROC) analysis was the primary analytic technique to address the central empirical issues of this study (Quinsey et al., 2006, pp. 49–60). This technique has become the most commonly used analytic procedure in recently published predictive validation studies of family violence risk assessment instruments (Heckert & Gondolf, 2004; Hilton & Harris, 2009; Hilton et al., 2008; Williams & Grant, 2006; Williams & Houghton, 2004). Moreover, it is the most appropriate when dichotomous family violence recidivism measures are used, as in this study. This analytic technique can be summarized briefly.

Violent recidivists accurately predicted to be violent represent "true positives" (i.e., sensitivity), and nonviolent recidivists accurately predicted to by nonviolent represent "true negatives" (i.e., specificity). False positives (nonviolent recidivists predicted to be violent) and false negatives (violent recidivists predicted to be nonviolent) constitute errors in prediction. A ROC curve is constructed by plotting the proportion of true positives against the proportion of false positives (i.e., 1-specificity) for each value of the prediction instrument (or predefined cutoffs on that instrument, e.g., quintiles or deciles).

The area under the curve (AUC) measures the accuracy of prediction. AUC will be .50 when the proportion of true positives equals the proportion of false positives across the scoring categories of the prediction instrument, meaning it does no better than chance in predicting recidivistic violence. Perfectly accurate prediction (predicted and actual outcomes are identical) would yield an AUC coefficient of 1.0, meaning the proportion of true positives is 1.0 and the proportion of false positives is 0 across all scoring categories of the risk assessment instrument. Neither extreme has been reported in the family violence risk assessment research literature. Rather, previous predictive validation studies have typically reported AUC coefficients ranging from about .60 to .80 (e.g., Hilton et al., 2008).

Of course, continuous family violence recidivism measures could also have been used. However, the distributions of those variables, like virtually all types of violent behavior, were positively skewed. Most cases were at 0, and the distribution of other cases dropped off quickly, with a very few extreme outliers. For example, the continuous distribution of total family violence offenses had a mean of 1.28, a standard deviation of 2.00, and a range of 0–19. Almost 80% of the cases was either not re-arrested or re-arrested only one or two times. Less than 1% was re-arrested 9–19 times. The Poisson-like distribution was even greater for the other family violence recidivism measures. Negative binomial regression is the most appropriate procedure for analyzing variables having these distributions. Such an analysis was done with the data used in this study, and the results paralleled those reported below (the results are available upon request from the author).

Results

Characteristics of perpetrators and descriptive statistics for the different types of family and household relationships, the DVSI-R total numeric risk scores, the perceived imminent risk ratings, and the five recidivism measures are presented in Table 1. The majority of the sample was male (70.55%), the largest ethnic group was Non-Latino White (47.16%), and the average age was 35. Intimate partners accounted for the majority of the family violence (64.62%). The mean DVSI-R total numeric risk score was 8.28, with the mean perceived imminent risk-to-victim being moderate (.92) and that for risk-to-others low to moderate (.59). Concerning the five measures of recidivism, about a third (33.80%) of the sample was re-arrested for new family violence offenses only, and less than 6% (5.53%) was re-arrested for violations or protective or restraining orders only. Almost 20% (19.09%) was re-arrested only for both, with overall family violence recidivism being 44.60%. Slightly more than 50% (52.14%) of the sample was re-arrested for some criminal offense, family or non-family related.

Cronbach's Alpha for the 11 items of the DVSI-R was .75, with the average inter-item covariance being .18. The range of item-total scale correlations extended from a low of .24 (children

Table 1
*Descriptive Statistics (Percent of Sample or Mean With Range in Parentheses) for the Connecticut Sample (*N = *3,569)*

Characteristics of perpetrators	
Gender (% male, *N* = 2,518)	70.55
Age (in years)	35 (18–81)
% Non-Latino White (*N* = 1,683)	47.16
% Black (*N* = 952)	26.67
% Latino (*N* = 687)	19.25
% Other or unknown (*N* = 247)	6.92
Household relationship types	
Intimate partner (%, *N* = 2,292)	64.62
Parent–child (%, *N* = 484)	13.64
Other types (%, *N* = 771)	21.74
Risk assessments	
DVSI-R total risk scores	8.28 (0–26)
Imminent risk-to-victim	.92 (0–2)
Imminent risk-to-others	.59 (0–2)
Dichotomous family violence recidivism measures	
New family violence offenses only (%, *N* = 872)	33.80
Violations of protective or restraining orders only (%, *N* = 100)	5.53
Both new and protective or restraining order violations only (5, *N* = 403)	19.09
Total family violence offenses only (%, *N* = 1,375)	44.60
Total all arrests combined (%, *N* = 1,861)	52.14

Note. The total all arrests category includes 486 offenders re-arrested for other non-family violence offenses. The percentage re-arrested for each of these categories is computed from sub-samples including those not re-arrested for anything and those re-arrested for specific violations (e.g., the sub-sample for new family violence offenses only includes the 1,708 cases not re-arrested for anything and the 872 offenders re-arrested only for such offenses, the sub-sample for the violations of protective or restraining orders only includes the 1,708 cases not re-arrested for anything and the 100 offenders re-arrested only for those violations, etc.).

present during any prior or current family violence incident) to a high of .72 (evidence of prior family violence assaults, threats, or arrests). Others have argued that such tests for internal consistency have "limited utility" for risk assessment instruments, with maximum predictive validity and limited redundancy among items being more relevant; however, inter-rater reliability is also important to ensure consistency in the assessment of family violence cases across independent evaluators (Kropp & Hart, 2000, p. 109). Establishing inter-rater reliability was an integral part of the training process. Specifically, case files for three separate family violence cases were used in training all Family Relations Counselors. An agreement rate of 80% was used to determine whether consistency among counselors was being achieved during the training sessions. No separate tests of inter-rater reliability post completion of the training sessions have yet been conducted.

Predictive Accuracy of the DVSI-R and Perceived Imminent Risk

The empirical results bearing on an initial assessment of the predictive accuracy of the DVSI-R, compared to the perceived imminent risk ratings are presented in Table 2. Consider the empirical relations between the DVSI-R total numeric risk scores and each measure of recidivism. Those scores were significantly associated with all recidivism measures, with the AUC coefficients (AUC = .621–.726) well within the range of those previously reported in the literature. Moreover, those coefficients were the strongest for violations of protective and restraining orders and the measure of such violations combined with new family violence offenses. The next strongest effects were on the overall measure of family violence recidivism and total rearrests. The weakest effect was for new family violence offenses only.

Now consider whether adding perceived imminent risk-to-victim or risk-to-other significantly added to the predictive accuracy of the DVSI-R total numeric risk scores. This analytic task was accomplished in a stepwise fashion by estimating logistic regression models for each measure of recidivism. The first step included the DVSI-R total numeric risk scores only, the second added the perceived imminent risk-to-victim, and the third added the perceived imminent risk-to-other. Predicted probabilities were saved at each step. These predicted probabilities were then used to calculate the AUC coefficients for each measure of recidivism (see Cleves, 2002, p. 304–309 for a discussion of this procedure in Stata).

As shown in Table 2, the addition of the perceived imminent risk measures did not significantly improved predictive accuracy across any of the five recidivism measures. However, an examination of the logistic regression results revealed that the perceived imminent risk-to-other had a statistically significant, independent estimated effect on total family violence offenses only (OR = 1.27, $z = 3.10$, $p = .00$) and total all offenses combined (OR = 1.28, $z = 3.36$, $p = .00$). The estimated effects of perceived imminent risk-to-victim were never statistically significant in any of the logistic regression models in which all the three risk assessment predictors were included simultaneously.

Moderating Effects of Perpetrator Characteristics and Household Relationship Types

It is already known that one of the research objectives was to determine whether the DVSI-R applies equally well across differ-

Table 2
AUC Coefficients, With 95% CI in Brackets, for the Family Violence Recidivism Measures and Total all Re-Arrests Combined for the DVSI-R, Imminent Risk-to-Victim, and Imminent Risk-to-Others

	DVSI-R	Risk-to-victim	Risk-to-others
New family violence offenses only	.621 [.598–.644]	.623 [.599–.646]	.630 [.607–.644]
Protective-restraining order violations only	.715 [.662–.768]	.728 [.676–.779]	.728 [.676–.779]
Both new offenses and protective-restraining order violations only	.726 [.699–.754]	.729 [.702–.757]	.731 [.703–.758]
Total family violence offenses only	.659 [.639–.678]	.661 [.642–.681]	.665 [.646–.685]
Total all offenses combined	.657 [.640–.676]	.659 [.641–.677]	.663 [.645–.681]

ent types of perpetrators and different types of family or household relationships involved in family violence. This objective was addressed analytically by testing the null hypothesis of no statistically significant difference between the AUC coefficients for males compared to females, younger compared to older offenders, Non-Latino Whites compared to ethnic minorities, and among the coefficients for the three family and household relationship types: intimate partner, parent–child, and other family and household relationships. Table 3 displays the results of the ROC analysis for perpetrator characteristics.

The key finding reported in Table 3 is that no evidence was found that these characteristics moderated the estimated effects of the DVSI-R on these recidivism measures. Specifically, tests for the statistical significance of the difference between the AUC coefficients failed to reject the null hypothesis, as indicated by the overlapping 95% Confidence Intervals for all comparisons, with one exception. The coefficient for females (AUC = .79) was significantly greater ($\chi^2(1) = 7.52$, $p < .01$) than the coefficient for males (AUC = .70) with regard to the joint measure of violations of protective or restraining orders and new family violence offenses only. Yet both coefficients were rather strong in magnitude.

Were the estimated effects of the DVSI-R on the behavioral recidivism measures moderated by type of family or household relationship? The answer to this question can be found in Table 4. No statistically significant differences between the AUC coefficients were found among the different types of family or household relationships for any of the behavioral recidivism measures, as indicated by the overlapping 95% Confidence Intervals, with one exception. The coefficient for parent–child (AUC = .83) was significantly greater ($\chi^2(1) = 6.78$, $p < .05$) than that for intimate partner (AUC = .70) concerning the joint measure of violations of protective or restraining orders and new family violence offenses only. However, once again, those coefficients were rather strong in magnitude.

Interpreting DVSI-R Risk Scores

Service providers and criminal justice practitioners are commonly challenged with interpreting risk scores the moment they have been derived from the application of a risk assessment instrument. One approach to meeting this challenge is to provide an interpretive framework that gives meaning to individual scores. Such a framework can be constructed by linking scoring categories of the risk assessment instrument to the recidivism rate within each of those categories (Hilton et al., 2004; Hilton et al., 2008). In short, a scoring category of risk is interpreted in terms of the likelihood of actual recidivism.

This approach is illustrated in Table 5. It shows the number of cases and the cumulative percentage distribution of those cases across DVSI-R total numeric risk scores subdivided into four approximately equal categories (quartiles). More importantly, Table 5 also shows the percentage of the sample re-arrested (i.e., the recidivism rate) for new family violence offenses only, violations of protective or restraining orders only, both new family violence offenses and violations of protective or restraining orders only, total family violence recidivism, and all the offenses combined during the 18-month follow-up period in each DVSI-R scoring category. The distributions clearly revealed the sharp increases in behavioral recidivism, irrespective of type, moving from one scoring category to the next on the DVSI-R. These categories could be equated with low, moderate, high, and very high levels of family violence risk.

Summary and Discussion

The predictive validity of the DVSI-R had been previously documented in a prospective study based on a sample of 14,970 assessments conducted between September of 2004 and May of 2006 in Connecticut (Williams & Grant, 2006). However, that research had two major limitations: (1) it was based on the developmental sample used to construct the instrument, and (2) the measure of recidivism itself—number of assessments conducted during the 6-month period (i.e., the number of times a perpetrator came to the attention of authorities) instead of re-arrests for repeated offending during that period.

This study addressed these limitations by using an independent, cross-validation sample and more refined measures of recidivism. Specifically, risk assessments were conducted on 3,569 defendants arrested on family violence charges across the State of Connecticut during February and March of 2007. Five refined recidivism measures were analyzed: re-arrests for new family violence offenses only, re-arrests for violations of protective or restraining orders only, a combined measure of such violations and new

Table 3
Testing for Differences in AUC Coefficients for the DVSI-R by Gender, Age, and Ethnicity

Demographic categories	Recidivism measures									
	NFVO		VPRO		BOTH		TOTFV		TOTA	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
Male	.61	[.580–.635]	.70	[.642–.762]	.70	[.663–.729]	.64	[.618–.665]	.64	[.621–.664]
Female	.63	[.588–.673]	.68	[.562–.802]	.79	[.730–.840]	.67	[.635–.708]	.66	[.630–.696]
Younger (18–35)	.62	[.593–.653]	.72	[.643–.797]	.72	[.680–.757]	.65	[.628–.680]	.65	[.626–.675]
Older (36–84)	.63	[.589–.662]	.70	[.624–.773]	.74	[.699–.779]	.67	[.641–.699]	.67	[.645–.698]
Non-Latino White	.64	[.570–.632]	.70	[.628–.778]	.73	[.692–.773]	.68	[.649–.705]	.67	[.644–.696]
Ethnic Minorities	.60	[.610–.678]	.72	[.644–.791]	.72	[.682–.758]	.64	[.614–.668]	.64	[.619–.669]

Note. *NFVO* new family violence offenses only, *VPRO* violations of protective or restraining orders only, *BOTH* new family violence offenses and violations or protective or restraining orders only, *TOTFV* total family violence offenses only, *TOTA* total all arrests combined.

Table 4
Testing for Differences in AUC Coefficients for the DVSI-R by Family and Household Relationship Types

	Recidivism measures									
	NFVO		VPRO		BOTH		TOTFV		TOTA	
Relationship types	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
Intimate partner	.62	[.588–.646]	.70	[.634–.758]	.71	[.680–.746]	.65	[.631–.679]	.66	[.634–.679]
Parent–child	.61	[.550–.674]	.63	[.344–.925]	.83	[.748–.911]	.65	[.595–.707]	.66	[.612–.710]
Other types	.63	[.587–.682]	.76	[.654–.871]	.72	[.656–.793]	.66	[.622–.706]	.65	[.609–.687]

Note. *NFVO* new family violence offenses only, *VPRO* violations of protective or restraining orders only, *BOTH* new family violence offenses and violations or protective or restraining orders only, *TOTFV* total family violence offenses only; *TOTA* total all arrests combined.

family violence offenses only, re-arrests for any type of family violence, and total re-arrests (all offenses combined, family or non-family) during an 18-month period after risk assessments were conducted.

Besides refining the measurement of recidivism, this study made two additional contributions to the research literature on family violence risk assessment. First, the analysis determined whether perpetrator characteristics or types of family or household relationships involved in family violence moderated the estimated effects of the DVSI-R on the five recidivism measures. In doing so, this research addressed a major limitation of virtually all previously published studies of family violence risk assessment—restricting the analysis to heterosexual intimate partners, with men as perpetrators and women as victims. All types of intimate partners, as statutorily defined by the State of Connecticut, were included in the sample analyzed (i.e., spouses and ex-spouses, adults having a child in common regardless of whether they have ever been married or resided together, previous or current unmarried couples who have resided together, adults in or previously in a dating relationship). Moreover, intimate partner violence included both men and women as perpetrators. The analyses also included family violence involving parents and children insofar as the perpetrators were 16 years of age or older, in addition to other types of family or household relationships involved in family violence. Besides gender, the moderating effects of age and ethnicity were analyzed as perpetrator characteristics. Granted, the composition of the sample of this study showed that 70.55% of the cases involved men as perpetrators, and 64.62% involved intimate partners (see Table 1). However, demonstrating that the DVSI-R maintained predictive accuracy across different perpetrator characteristics and different types of family or household relationships involved in family violence was vital if risk assessments are required for all family violence incidents resulting in an arrest within a population, which is the case in Connecticut and increasingly elsewhere.

Second, this research empirically examined whether incorporating a mechanism for clinically overriding the DVSI-R total numeric risk scores (i.e., the perceived imminent risk-to-victim and perceived imminent risk-to-others) corresponded with the prediction of family violence recidivism yielded by those scores.

Findings supported the predictive accuracy of the DVSI-R (see Tables 2, 3, 4). The AUC coefficients representing the estimated effects of the total numeric risk scores of that instrument on the behavioral recidivism measures showed a consistent pattern. They were the strongest for re-arrests for violations of protective or restraining orders only and the joint measure of those violations combined with new family violence offenses only, followed by the overall measure of family violence recidivism and total re-arrests. The weakest estimated effects were for new family violence offenses only. Moreover, this pattern persisted across categories of gender, age, ethnicity, and family or household relationship types. All coefficients were in the range of those previously reported in the literature and tended to be moderate to strong in magnitude. The implication is that the assessment of family violence risk using the DVSI-R is sensitive to actual family violence re-offending, and that this risk assessment instrument differentiates between specific forms of family violence re-offending.

Further, the estimated effects of the DVSI-R total numeric risk scores on total re-arrests (all violations and offenses combined, family and non-family) were moderately strong in magnitude. Admittedly, this finding could be the result of total re-arrests being dominated by cases of family violence since 73.89% of the 1,861 perpetrators re-arrested during the 18-month follow-up period were charged with some form of family violence crime. Nonetheless, although Family Relations Counselors in Connecticut administer the DVSI-R only for perpetrators arrested on family violence

Table 5
Interpreting DVSI-R Risk Scores in Terms of Recidivism (% in Each Refined Offending Category)

DVSI-R risk scores	Category (*N*)	Cumulative percent	NFVO	VPRO	BOTH	TOTFV	TOTA
0–4	1 (1,060)	35.13	26.60	2.30	7.10	29.02	35.85
5–8	2 (884)	61.14	34.03	4.33	16.29	43.04	50.00
9–12	3 (743)	80.81	36.91	6.10	22.54	48.48	56.53
13–28	4 (807)	100.00	50.11	15.50	41.28	65.41	71.62

Note. *NFVO* new family violence offenses only, *VPRO* violations of protective or restraining orders only, *BOTH* new family violence offenses and violations or protective or restraining orders only, *TOTFV* total family violence offenses only, *TOTA* total all arrests combined.

charges, this finding is important for evaluators in other jurisdictions required to administer family violence risk assessments on all the arrested perpetrators, irrespective of whether they have been charged with a family violence crime or not. The DVSI-R still maintained a moderately strong estimated effect on re-arrests for all criminal offenses combined.

With the exception of gender differences in the prediction of re-arrests for violations of protective or restraining orders coupled with new family violence offenses only, no evidence was found that the DVSI-R differentially predicts recidivism depending on the characteristics of the perpetrator. In addition, with one exception (parent–child and the joint measure of court order violations and new family violence offenses only), no evidence was found of any differential prediction based on the type of family or household relationship involved in the family violence.

These empirical results suggest the DVSI-R is a relatively robust family violence risk assessment instrument, having utility across perpetrator characteristics and types of family or household relationships involved in family violence. This finding is important since states across the country have broadened the types of relationships that qualify for family violence crimes, as noted in the introduction to this article. Nonetheless, future research can refine such analyses further by considering other characteristics of perpetrators, characteristics of evaluators, variations in criminal charges, or more detailed sub-categories of relationship types, including same-sex versus heterosexual relationships involving intimate partners or other family and household members (e.g., roommates, siblings, cousins, etc.). This research was at least an initial analysis of some potential moderators of the empirical relation between the DVSI-R and family violence recidivism.

The DVSI-R also included perceived imminent risk ratings whereby evaluators provided their clinical judgment of the imminent risk of future violence (within the next 6 months) to the victim or other persons known to the victim and/or perpetrator. The empirical results showed that adding perceived imminent risk-to-victim and risk-to-others did not significantly improved predictive accuracy beyond the DVSI-R total numeric risk scores. However, perceived imminent risk-to-other had an independent and statistically significant estimated effect on total family violence offenses and total all offenses combined in the logistic regression analysis, suggesting it is an important added component to the protocol for assessing the risk of family violence. That said, structured clinical judgments overall may be less relevant for an empirically based risk assessment instrument, that is, one representing numeric scoring and validated through a prospective study (see Hilton et al., 2008 and Hilton et al., 2010, pp. 69–106 for detailed analysis of this issue).

Predictive accuracy is only one issue concerning these perceived imminent risk ratings. Providing evaluators with a mechanism for supplementing numeric risk scores would be beneficial when those scores clearly misrepresent behavioral risk, as in the example provided previously. After all, unless an actuarial risk assessment instrument captures an exhaustive and exclusive domain of risk factors for family violence recidivism, and those risk factors are perfectly measured, prediction errors will occur. A mechanism such as the perceived imminent risk ratings provides a possible safeguard for prediction errors. This mechanism also avoids disempowering evaluators and thus achieving their support for a structured and objective risk assessment protocol. Given the findings reported here, however, perceived imminent risk ratings should not supersede numeric risk scores and should be used only as supplemental assessments with close and careful monitoring by authoritative supervision.

Drawing from previous research on the predictive validity of family violence risk instruments, a framework was presented to interpret DVSI-R total risk scores. The purpose was to give them meaning by linking assessed risk to actual risk, meaning the observed likelihood of family violence recidivism within the scoring categories of the DVSI-R. Of course, the framework should be reconstructed if the DVSI-R is used within a different population, if the DVSI-R is further modified, or if a different risk assessment instrument goes online. Moreover, recidivism should be monitored to determine whether additional adjustments to the framework are necessary to compensate for temporal variation in patterns of family violence re-offending. All such matters aside, the interpretive framework and the accumulating research evidence offering at least moderate support for the predictive validity of the DVSI-R was offered with the hope that it will promote more systematic and rigorous assessments of family violence perpetrators, with those assessments linked to more effective supervision and treatment.

References

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory-Revised. Toronto*: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct.* Newark, NJ: LexisNexis.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI).* Toronto: Multi-Health Systems.

Brennen, T., Dieterick, W., & Ehret, B. (2009). Evaluating the predictive validity of the Compas risk and needs assessment system. *Criminal Justice and Behavior, 36,* 21–40. doi:10.1177/0093854808326545

Campbell, J. C. (2007). Prediction of homicide of and by battered women. In J. C. Campbell (Ed.), *Assessing dangerousness* (2nd ed., pp. 85–104). New York: Springer.

Cleves, M. A. (2002). From the help desk: Comparing areas under receiver operating characteristic curves from two or more probit or logit models. *The Stata Journal, 2,* 301–313. Retrieved from http://www.statajournal.com/article.html?article=st0023

Dutton, D. G. (2006). *Rethinking domestic violence.* Vancouver, BC: UBC Press.

Dutton, D. G., & Kropp, P. R. (2000). A review of domestic violence risk instruments. *Trauma, Violence, & Abuse, 1,* 171–181. doi:10.1177/1524838000001002004

Goodman, L. A., Dutton, M. A., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers: Use of the Danger Assessment Scale in the criminal justice system. *Journal of Interpersonal Violence, 15,* 63–74. doi:10.1177/088626000015001005

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

Heckert, D. A., & Gondolf, E. W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence, 19,* 778–800. doi:10.1177/0886260504265619

Hilton, N. Z., Carter, A. M., Harris, G. T., & Sharpe, A. J. (2008). Does using non-numerical terms to describe risk aid violence risk communication? Clinician preference, agreement, and decision-making. *Journal of Interpersonal Violence, 23,* 171–188. doi:10.1177/0886260507309337

Hilton, N. Z., & Harris, G. T. (2005). Predicting recidivism among serious wife assaulters: A critical review and implications for policy

and practice. *Trauma, Violence, & Abuse, 6*, 3–23. doi:10.1177/1524838004272463

Hilton, N. Z., & Harris, G. T. (2009). How nonrecidivism affects predictive accuracy: Evidence from a cross-validation of the Ontario Domestic Assault Risk Assessment (ODARA). *Journal of Interpersonal Violence, 24*, 326–337. doi:10.1177/0886 260508316478

Hilton, N. Z., Harris, G. T., & Rice, M. R. (2010). *Risk assessment for domestically violent men: Tools for criminal justice, offender intervention, and victim services.* Washington, DC: American Psychological Association. doi:10.1037/12066-000

Hirschel, D., Buzawa, E., Pattavina, A., Faggiani, D., & Reuland, M. (2007). *Explaining the prevalence, context, and consequences of dual arrests in intimate partner cases.* Final report to the U.S. Department of Justice, Washington, DC.

Kropp, P. R., & Hart, S. D. (2000). The Spouse Assault Risk Assessment Guide (SARA): Reliability and validity in adult male offenders. *Law and Human Behavior, 24*, 101–118. doi:10.1023/A:1005430904495

National Research Council. (2004). *Advancing the federal research agenda on violence against women.* Washington, DC: The National Academic Press.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, M. A. (2006). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association. doi:10.1037/11367-000

Simpson, S. S., Bouffard, L. A., Garner, J., & Hickman, L. (2006). The influence of legal reform on the probability of arrest in domestic violence cases. *Justice Quarterly, 23*, 297–316. doi:10.1080/07418820600869087

Williams, K. R. (2008). Domestic Violence Screening Instrument (DVSI). In B. L. Cutler (Ed.), *Encyclopedia of psychology and law* (pp. 240–242). Newbury Park, CA: Sage Publications.

Williams, K. R., & Grant, S. R. (2006). Empirically examining the risk of intimate partner violence: The revised Domestic Violence Screening Instrument (DVSI-R). *Public Health Reports, 121*, 400–408. Retrieved from http://www.jstor.org/stable/20056981

Williams, K. R., & Houghton, A. (2004). Assessing the risk of domestic violence re-offending: A validation study. *Law and Human Behavior, 28*, 437–455. doi:10.1023/B:LAHU.0000039334.59297.f0

Williams, K. R., Tuthill, L., & Lio, S. (2008). A portrait of juvenile offending in the United States. In R. D. Hoge, N. G. Guerra, & P. Boxer (Eds.), *Treating the juvenile offender* (pp. 15–32). New York: The Guilford Press.