Intimate Partner Violence Risk Assessment Validation Study
NIJ 2000WTVX0011
Final Report: March 28, 2005

PROJECT PERSONNEL

Johns Hopkins University
Jacquelyn C. Campbell, PhD, RN, Principal Investigator
Daniel Webster, ScD, Co-Investigator
Patricia Mahoney, MA, Project Director

Safe Horizon, New York
Chris O'Sullivan, PhD, Co-Principal Investigator
Melissa White, MSW, Project Director; Johanne Eliacin, MA, Project Director

Justice Research Center, California
Janice Roehl, PhD, Co-Principal Investigator
Kristen Guertin, Project Director
Kate Semple, Project Director

OVERVIEW

The major purpose of this multi-site field test, funded by the National Institute of Justice, was to assess the predictive accuracy of several methods of assessing risk of repeat assault or potential lethality in domestic violence cases. The methods tested were chosen because they are currently being used by agencies and service providers around the country. One of the instruments, the Danger Assessment (DA), and the threat assessment method, DV-MOSAIC, include questions or domains of inquiry designed to diagnose risk of lethal violence and extreme dangerousness. Two of the instruments tested, the Domestic Violence Screening Instrument (DVSI) and Kingston Screening Instrument for Domestic Violence (K-SID) were designed to diagnose risk of repeat assault at any level. We also tested a number of other questions, drawn from other instruments, that research shows may be diagnostic of dangerousness. For those instruments and the assessment method that yield an overall rating, we evaluated the correlation of those ratings with subsequent re-assault, stalking, serious injury or death. In assessing the predictive accuracy of the three instruments and one threat assessment method, we took into account protective actions taken, which may mitigate risk. The major goal of the field test was:

To test the predictive accuracy of two instruments which assess risk of repeat violence (the DVSI and K-SID) in intimate relationships, and one instrument (the Danger Assessment) and one Threat Assessment Method (DV-MOSAIC) assessing lethal or near lethal risk in violent intimate partner relationships.

We also aimed to determine the strength of prediction of the individual risk factors found in these instruments/method, other assessment instruments, and the DV literature, and in a more limited fashion, to evaluate the impact of risk assessment on victims' self-perceived risk and self-protective behaviors. In this report, we will report on the results of the major purpose of the study. Additional analyses to address the secondary purposes are not complete and will be disseminated in future manuscripts.

A. BACKGROUND AND SIGNIFICANCE

PURPOSE AND OBJECTIVES

Women in the United States are murdered by an intimate partner (married and non-married) or former partner more often than by an acquaintance or stranger. In fact, at least 30% of American women who are killed are murdered by an intimate partner or ex-partner according to the Supplemental Homicide Records (SHR; NJC, 2000). When hand counts correct for the misclassifications in the SHR (Langford, 1992), the percentage increases to 40 – 50% (Campbell et al., 2003a; 2003b). In cases of intimate partner homicide, the victim or perpetrator or both have usually had contact with criminal justice, victim assistance, and/or health agencies (Sharps et al., 2001). The agencies will hold a fatality review and ask themselves, "Could we have known? What were the signs?" The public and press will ask, "Why wasn't more done to protect her?" In hindsight, there usually was a sign. Often, there was a failure to read that sign or to act on it. Sometimes it is difficult to convince the victim that the risk was serious enough to warrant drastic and penalizing alterations to her lifestyle to stay alive.

Galina Komar, who had been assaulted and was being stalked by her ex-boyfriend, told a judge that he would kill her. The judge released him from jail on time served because there were no new charges. When her former partner tracked Ms. Komar down at her job at a car sales lot and shot her and himself, the judge said the case was indistinguishable from the hundreds of other misdemeanor assault cases he presided over that did not end in murder. Yet research shows that victims can be more accurate than others in identifying the risk of severe and potentially lethal violence (Goodman, Dutton et al., 2000; Gondolf & Heckert, 2002; Weisz, Tolman & Saunders, 2000). Ms. Komar was participating in a project of the DA's office that provided high- risk victims with cell phones connected to the police, but she had not brought the phone to work that day. In any case, it is unlikely that the police could have responded quickly enough to prevent the murder-suicide that occurred within minutes of the confrontation.

In another case, a man who was participating in a batterer program was despondent over the prospect of spending Christmas alone, separated from his wife and children by court order. The other men in the class invited him over and the instructor gave the man his cell phone number. His wife had declined the services of a police program that offers counseling and safety planning to victims in cases identified as high risk because of repeated 911 calls. Shortly after the holiday, the man shot his wife and himself. The batterer program changed its protocol to include depression as a danger sign. The victim assistance program wondered what they could have done to encourage the victim to avail herself of their services, to take her situation as seriously as they did.

In yet another fatality review, the police officer who had responded to the most recent call said that, although the crime scene looked like all the other domestic violence calls he had handled when there was minor violence and no injuries, this time, when he interviewed the victim, the "hair on the back of his neck stood up." He didn't know why.

Finally, in another case, a man was evaluated for suicidal intent by a psychiatrist on call from an emergency department. Although well trained in traditional suicidality risk, this mental health professional failed to assess for and take into account the salience of domestic violence in the

man's marriage, his wife's recent separation, as well as his destruction of her property and his access to guns. The psychiatrist decided that the man was not immediately suicidal and let him go. Within hours of his release, the man killed his wife and then himself. Their eight year old son heard the shots and found the bodies.

Using data from a recently completed study of intimate partner femicide in 11 cities, Sharps, Campbell and associates (2001) found that although only 5% had gone to a domestic violence shelter, 74% of the 239 murdered women and 88% of victims of attempted femicide had been seen in emergency departments for some ailment during the year before the incident. Thirty-two percent of the women had sought help at hospital emergency departments, hospital inpatient units or ambulatory care settings for injuries specifically resulting from the abuse. Just over half (51%) of the perpetrators had been previously arrested by police: 44% were arrested for prior domestic abuse incidents, 37% for other violent crimes and 58% for other non-violent crimes. Thus, the anecdotes above are not atypical and dramatize the need for law enforcement, the courts, victim assistance programs and hospital emergency departments to have valid and systematic means of evaluating domestic violence cases and identifying those most likely to escalate to lethality. In addition, in spite of the evidence that abused women's perception of high risk is often accurate, the 11- city femicide study found that only about half of the women who were victims of actual or attempted intimate partner homicide accurately assessed their risk (Campbell et al., 2003b). As in other studies of abused women, they underestimated the threat of lethality or diminish the severity of the violence being perpetrated against them. This minimization has been interpreted as an adaptive coping mechanism that allows victims to continue to invest in the relationship and reduce the distressing symptoms of trauma (e.g., Campbell et al., 1998). Weisz also sees it as a survival strategy, when leaving is likely to trigger lethal violence (Weisz, Tolman & Saunders, 2000). The history of the development of the coordinated community response to domestic violence reveals other reasons for validating formal instruments for evaluating risk.

Project Need

Over the past two decades, there has been a dramatic transformation in the response to intimate partner violence across all sectors of society, including the criminal justice system, social services, health care and public opinion. Enhancement of the response in the criminal justice system includes a change in law enforcement officers' treatment of domestic violence calls, such policies as mandatory arrest and prosecution, training for law enforcement and judges on the dynamics of domestic violence, increased penalties, and special domestic violence courts. In social services, the last twenty years have seen the creation of hotlines and emergency shelters for battered women and their children across the nation, creation of advocacy programs in family and criminal courts, and counseling programs in virtually every community. In a more recent development, child welfare programs are beginning to take domestic violence into account and to offer services to battered mothers as an integral component of child safety. In health care settings, emergency departments and prenatal settings increasingly screen for domestic violence and many hospitals have advocacy and counseling programs for victims of domestic violence. Public awareness of domestic violence as a crime, not a private family matter, has also altered the landscape and increased the demand for services (Kline, Campbell, Soler & Ghez, 1997).

The increased use of criminal justice remedies for intimate partner violence has necessitated a sort of triage in case processing by law enforcement, prosecutors and the courts. The demand for emergency shelter exceeds supply (Crowell & Burgess, 1996), and the number of cases entering

the courts prohibits incarceration of every individual convicted of a misdemeanor for assaulting or harassing an intimate partner. Determining the seriousness of a particular case and risk of escalation is necessary not only for the obvious reason of allocating resources. It is also important to tailor the response to the level of dangerousness, to make the response appropriate, in order to avoid violating the civil rights of offenders and to avoid unnecessarily disrupting the lives of victims and their children. An overzealous response can discourage victims from accessing services and using the criminal justice system. At the same time, it may be necessary in some cases for service providers and law enforcement to try to convince victims that they are in a high-risk situation when they believe the abuser will not hurt them again. Convincing a victim that it is necessary to leave her home and friends and take her children out of school to enter an emergency shelter or relocate requires her belief that she has no alternative, that staying home is really not safe. Convincing a woman to cooperate in having her partner tried on a felony charge and possibly incarcerated, depriving the children of a father, the family of a source of financial support, and treating the person she is emotionally attached to as a criminal may require persuasive evidence. Several experts in this field have maintained that the process of risk assessment can help victims come to a more realistic appraisal of the danger in their situation (Humphreys & Campbell, 2004; Campbell, 1995; de Becker, 2000).

In part, the attempt to predict which batterers are most dangerous and which victims are most at risk stems from the apparent predictability of this crime. It is typically a repetitive crime, with the same perpetrator and victim. Furthermore, there is a wide spread belief (not fully supported by research) that the violence usually escalates over the course of the relationship. There is evidence that the risk of severe violence and murder may increase when the victim attempts to end the relationship, and if she separates from the abuser (Daly & Wilson, 1993; Campbell et al., 2003), although no comparison has been made with abused women who stay in the relationship. Clearly, however, only a small percentage of violent intimate partnerships end in murder. Because, however, the perpetrator and victim may already be known to the system, and observations have been made of common characteristics of the cases that do end in homicide, it seems incumbent on service providers to attempt to identify the high-risk cases and intervene most strenuously in those cases. For example, police might opt for arrest, prosecutors might advocate for incarceration over alternative sentencing, courts might provide expedited prosecution, and shelters might give priority to high-risk cases even though demand exceeds supply, and probation departments might be more vigilant and check in more frequently if we could identify the most dangerous offenders.

Responding to this need, programs and agencies dealing with batterers and victims, such as police programs, victim assistance agencies and probation departments, have adopted a number of mechanisms to identify high-risk cases. The mechanisms include checklists, clinical interviews or practitioners' intuitions, and formal assessment instruments (Campbell, 1995). There has been a great deal of interest among law enforcement in the MOSAIC 20 and its later version DV-MOSAIC, a computer-assisted threat assessment method developed by Gavin de Becker & Associates. The MOSAIC 20 has been purchased by POST, California's main police standards and training organization, which gives it to all law enforcement officers who attend DV training workshops. (We do not know the extent of its usage, however.) MOSAIC 20 has since been succeeded by DV-MOSAIC, which is more an investigation guide and training method than a risk assessment questionnaire. Probation departments in Colorado use the Domestic Violence Screening Instrument (DVSI), and probation departments in Connecticut used the Kingston Screening Instrument for Domestic Violence (K-SID; Gelles).

Minnesota law requires a pre-sentence domestic violence investigation conducted by the appropriate corrections agency before a perpetrator of DV is sentenced by a judge. The report is to include: "a recommendation on any limitations on contact with the victim and other measures to ensure the victim's safety" (MS 609.2244). Some counties have introduced forms of risk or dangerousness assessment to meet this requirement. For example, Hennepin County (Minneapolis) uses the DVSI; if an offender scores high on that brief instrument, the investigators administer the SARA (Spousal Abuse Risk Assessment). Police departments, shelters, and hospital-based domestic violence programs have adopted the Danger Assessment (Campbell, 1995). The Navy has developed a risk assessment inventory, and the Duluth program has developed the PSI to assess system safety as well as individual risk. Most commonly, however, practitioners assess risk "clinically," relying on beliefs or knowledge, experience and intuition, or they use an ad hoc scale or in the case of law enforcement, arrest records or repeat calls to the same location. Some entire states (such as Colorado) and provinces (such as Ontario) in Canada have mandated risk assessment in all domestic violence cases in the criminal justice system.

None of the instruments in use has been rigorously independently or comparatively validated using a prospective design. The lack of such validation has discouraged some agencies from adopting an instrument, although they are often under pressure to have some method of identifying the cases most likely to escalate to homicide.

Purpose and Utility to Criminal Justice System

The primary purpose of this research project was to conduct a field test of some of the instruments and methods developed specifically to assess risk of repeat intimate partner violence and severe or lethal violence in various settings. This prospective study produced quantitative results showing the comparative predictive validity of instruments/assessment methods currently in use by law enforcement, probation and victim assistance programs. The data also suggest the degree to which agencies should (or should not) rely upon such assessments; and their validity for different levels of risk and different populations. Overall, the results should help criminal justice agencies and their cooperating partners decide whether to use a formal risk assessment instrument or method and, if so, which to adopt in different settings and what role formal assessment should play. A limitation of the study is that we will not, in a single study, be able to address all the important questions regarding the impact of risk assessment in the field.

Future studies will have to determine the impact of risk assessment (possibly with the most predictive instrument or method identified in this experiment) on agencies' decision making and victims' perceptions and actions, utilizing a control group that does not receive a formal risk evaluation and a comparison group that receives a clinical assessment. Also, the feasibility, utility and impact of risk assessment need to be investigated in different settings, relying upon different information sources. Although this study did recruit participants from many types of settings, the primary information source was the victim, and the risk assessments were administered by researchers rather than law enforcement personnel or practitioners. Actual field use and implementation could present a different picture.

Report Organization

A one-year planning study (O'Sullivan, Campbell & Roehl, 1998, NIJ Grant 98-WT-VX-0019) was carried out to resolve issues of victim safety and such concerns as the duty to warn victims at risk;

methodological issues, such as populations to be tested, the sample size necessary to capture all levels of risk that present to different agencies, and the appropriate follow-up interval and implementatiion issues, including identifying field sites and reaching agreements with them about contacting victims seeking assistance, and negotiating an agreement with Gavin de Becker & Associates on the use of DV-MOSAIC.

The findings and outcomes of the planning study are presented as the relevant parts of the methods section rather than a separate section. The introductory section of the report is a comprehensive review of the state of the science in assessment of risk of intimate partner homicide and re-assault, with focus on the extant reliability and validity data of the risk assessment instruments and methods in use, one of the outcomes of the planning study. The literature review is broad, covering the general topic of assessing risk, specific risk factors in domestic violence, and the state of the science, including development of domestic violence risk instruments and risk assessment methods. The next section describes the methodology of the study, including staffing, the procedures followed in each of the settings, description of the analysis plan, human subjects issues, and limitations and obstacles. The rest of the report presents the study results, and the final section discusses the findings, limitations of the study, and recommendations for policy and future research.

B. PREDICTING DANGEROUSNESS IN INTIMATE PARTNER VIOLENCE: THE STATE OF THE SCIENCE

Approaches to Assessing Risk of Violence

Clinical vs. Actuarial Assessment

Predicting dangerousness or repeat violence in cases of intimate partner violence (IPV) is of increasingly intense interest to both the advocacy community and the criminal justice system. In the mental health field, there has been a longer history of interest in predicting dangerousness of individuals and more methods have been developed and tested, as part of a tradition of diagnosis and prognostication through empirically validated testing. Therefore, we turn first to psychiatry for models of assessing risk of violence. There has been a long and vigorous debate among mental health researchers about whether formal instruments – the "actuarial method" – are more valid than clinical judgment (e.g. Quinsey et al., 1998; Litwack & Schlesinger, 1999). An actuarial method can be defined an instrument that provides weightings and empirically based scores that relate predictor variables selected from measured associations with criterion variables in representative samples (Hilton, Harris, Rice, Lang, Cormier & Lines, 2004; Webster, Harris, Ricke, Cormier, & Quinsey, 1994). In assessing violence risk, an actuarial instrument is one that has been formally and independently tested and shown actually to *predict* violent outcomes.

Litwack and Schlesinger (1998) maintain that the current state of the science does not support the accuracy of actuarial methods, while Quinsey (Quinsey, Harris, et al., 1998) and others maintain that actuarial methods are the only defensible approach. A recent meta-analysis of 95 studies of re-offending among sexual offenders (Hanson & Morton-Bourgon, 2004) first of all demonstrated how much further along the path to actuarial methods the sexual assault field is than the intimate partner violence field, given the sheer number of studies available for analysis. The findings also supported the use of actuarial risk assessment instruments over "unguided clinical opinion" with "empirically guided professional judgments" intermediate between the two. Thus, a dichotomy is

probably overly simplistic, since the best approach is to gather as much information (clinical and actuarial) from as many sources as possible in the time available and under the extant circumstances (Litwack, Stuart ,et al., 1993; Litwack & Schlesinger, 1998; Monahan, 1996; Mulvey & Lidz, 1995). The ideal would be a well-validated instrument specific to the type of violence being assessed in the hands of a clinician who is expert in that type of violence by virtue of training and experience. Kropp (2004) suggests the terminology "structured professional judgment" to describe this kind of approach. This approach also would include guidelines for how to collect the information and communicating the results of any assessment.

The Hanson and Morton-Bourgon (2004) meta analysis of sexual assault offenders used the d. statistic, a standardized mean difference between recidivists and non-recidivists. This test showed no significant differences among the six major sexual re-offending measures, although their predictive accuracy ranged from d. = .52 -.77. The meta-analysis also showed that general criminal recidivism actuarial measures did as well as the sexual assault specific instruments in predicting sexual reassault and were better at predicting general (any) criminal recidivism, the more common reoffense of this type of perpetrator. Little of this kind of work has been done in the domestic violence field, nor are there a sufficient number of data sets upon which to conduct a meta-analysis.

Comparison between Assessing Risk of Community Violence Among Mentally Ill and Assessing Risk of Intimate Partner Violence

Even accepting Quinsey's argument that the actuarial approach is preferable, the adequacy of actuarial assessment depends on the development of the science in the specific area of violence under consideration and type of perpetrator involved. In Quinsey's major area of assessment, sexual assault perpetrators, there is time to conduct assessment of the perpetrator, instrumentation (phallometric assessment) and actuarial science development, permitting the kind of composite evaluation that best predicts re-offending (Quinsey et al., 1998). The most research on assessing the risk of future violence has been conducted in the areas of sexual assault, general violent behavior, and community violence by the diagnosed mentally ill. The applicability of these developments to intimate partner violence is not clear, nor is it clear how much of a risk psychopathology in general and psychopathy specifically is for perpetration of either domestic violence or intimate partner homicide.. Sexual assault perpetrators and batterers have not been compared (with the exception of O'Sullivan's [1998] article on gang rapists and batterers, which found important differences between the two groups).[1]

Most studies have not found consistent patterns of psychopathology among batterers, leading several researchers to propose typologies of batterers with differential patterns of mental illness (e.g. Holtzworth-Munroe & Stuart, 1994). Others propose that domestic violence has been a socially acceptable expression of male dominance, and therefore, while a small percentage may be mentally ill, the majority of batterers are merely suffering from socialization to masculinity and

[1] We recognize that batterers frequently abuse their partners sexually, but we are drawing a distinction here between men whose primary crime is sexual assault of strangers, acquaintances and perhaps intimate partners, and men whose primary crime is abuse of an intimate partner, including sexual abuse.

patriarchal enculturation. In sum, borrowing instruments from mental health assessments of risk of violence may therefore be inappropriate.

State of the Science of Assessment of Risk of Intimate Partner Violence

In the intimate partner violence field, checklists developed to aid practitioners' expert judgment (e.g. Hart, 1988) are widely used. More recently, several instruments have been developed to assist domestic violence service providers, law enforcement and the courts in determining the cases most likely to escalate to severe or lethal violence, and a number of these instruments and *ad hoc* scales are in use. Several of the instruments have recently published predictive validity support but none have been widely validated in independent studies conducted by other researchers than the instrument developers. Therefore, none can be defined as an independently validated actuarial method, the kind of instrument sought by the field. Furthermore, in contrast with mental health practitioners conducting clinical assessment, police and probation officers, victim advocates and health care providers who come into contact with batterers and victims of intimate partner violence are not primarily and specifically trained in clinical assessment. Their expertise and judgments may be equally valid, but come from experience and may vary widely according to the individual's interest in and understanding of domestic violence.

Even to the extent that there has been testing to validate individual instruments, there have been methodological problems because of the low base rate of the offense in question, especially if the criterion is re-arrest, or, even more so, homicide. The low rate of intimate partner homicide makes it especially statistically problematic to predict (Monahan, 1984).

The lack of validated instruments and the question of clinical vs. actuarial method are but two of the issues that are problematic in the current state of the science regarding risk assessment in intimate partner violence cases. A third is the frequent lack of clarity regarding the type of prediction required for the situation. Some of the instruments/threat assessment methods were created expressly to predict lethality or near lethality in domestic violence situations, most notably the MOSAIC-20 (Gavin de Becker & Associates) and the Danger Assessment (DA; Campbell, 1986; 1995; 2003). Others were designed to predict reassault, for instance the SARA (Spousal Abuse Risk Assessment; Kropp et al., 1995; Kropp & Hart, 2000), the DVI and the K-SID (Kingston-Screening Inventory for Domestic Violence; Gelles, 1995). Others attempt to predict both re-assault and lethality (Navy Risk Assessment and PSI).

The instruments and methods also vary in regard to the primary source of information (perpetrator, victim and/or other sources, such as criminal records), and which system the information is primarily intended to benefit, such as victim assistance programs, probation, etc. (Roehl & Guertin, 2000). Roehl and Guertin (1998, 2000) provided a comprehensive overview of these issues in relationship to the major instruments and methods that are currently in use.

Finally, there are unique ethical and empirical issues in assessing risk of intimate partner violence. A unique aspect of prediction in IPV is that a particular individual victim is the object of concern, rather than the population at large addressed in prediction of sexual assault and mental patient violence. Knowing the identity of the potential victim makes it possible, and therefore incumbent on service providers, to consider her (or his) safety as paramount. Furthermore, the known potential target will have her (or his) own perception of the dangerousness of the perpetrator, a prediction that may be more accurate than any instrument or clinician. The intended victim may have input

into the process of risk assessment, and may need to know the results of and system response to any formal risk assessment.

This unique situation creates the need for research on domestic violence risk assessment to take into account the victim's actions as a result of risk assessment and any proactive/preventive response on the part of the victim or criminal justice system. These actions will compromise the ability of the research to detect the accuracy in predicting recurrence of violence, especially lethal or severe reassault. (This issue is addressed in the proposed methodology.)

Risk Factors from Research

Intimate Partner Homicide - Homicide/femicide Rates

American women are most often killed by a husband or lover, or ex-husband or ex-lover (Mercy et al., 1989; Bailey et al., 1997; Bachman et al., 1995). Thus, intimate partner homicide is the largest category of murders of women, or femicide, accounting for approximately 30-40% of murders of women according to the official counts based on the Supplemental Homicide Reports (SHR; National Institute of Justice, 2000). Since the SHR misclassifies as many as 13% of intimate partner homicides of women as non-intimate partner, this percentage is undoubtedly an underestimate (Langford, Isaac & Kabat, 1998). The recent 11- city intimate partner femicide study (Campbell et al., 2003) found 19.3% of those intimate partner homicides of women to be perpetrated by an ex-boyfriend, a category of perpetrator not specifically accounted for by the SHR. A recent analysis of the data on homicides of women in 2001 found that husbands and intimates perpetrated in 51% of cases (Brock, 2003). In contrast to homicides of women, homicides by intimate partners account for a relatively small proportion of murders of men in the US, approximately 5-8% in 2000 (Lattimore et al., 1997; National Institute of Justice, 2000).

In recent decades in the US, intimate partner homicide rates overall have declined (Dugan, Nagin & Rosenfeld, 2003; Rosenfeld, 1997). One possible explanation for this trend is that the concurrent lower marriage rates in recent decades have decreased women's exposure to legally sanctioned spouses. However, while there has been a decrease in murders of married women by their husbands, there has been an increase in murders of non-married women by their boyfriends (Rosenfeld, 1997). Another reason for the decline appears to be lowered handgun availability, at least in some cities (National Institute of Justice, 1997; Wilt, Illman, & Brodyfield, 1995).

The largest decreases in intimate partner homicide have been for male victims. Consequently, the *proportion* of male homicides by female intimate partners has decreased and the proportion of femicides by male intimate partners has increased. From 1976 to 1996, the percentage of intimate partner homicides with female victims increased from 54% to 70% (National Institute of Justice, 1997; Zawitz, 1994). The decrease in the number of men killed by female partners coincided with the development of services for battered women and the enhancement of the criminal justice response. A connection has been made between the increased alternatives and protections available to battered women, such as the reduction of barriers to leaving violent relationships, legal sanctions and shelter resources and the decline in the male intimate partner homicide rate: women are able to secure safety from violence rather than kill an abusive partner (Browne et al., 1998; Dugan, Nagin & Rosenfeld, 2003; Rosenfeld, 1997).

Factors Associated with Intimate Partner Homicide and Femicide

Battering

Abuse is most often a precursor of intimate partner homicides, whether the victim is the male or female partner. The majority (67-75%) of intimate partner homicides involve battering of the female by the male intimate, no matter which partner is killed (Bailey et al., 1997; Campbell, 1992; Campbell et al., 2003; McFarlane et al., 1999; Mercy et al., 1989; Moracco et al., 1998; Pataki, 1998; Websdale, 1999). Two earlier American studies in different jurisdictions documented that two-thirds of the intimate partner femicide cases had a documented history of battering of the female partner (Moracco et al., 1998; Campbell, 1992). The recent 11-city study found that 72% of the intimate partner femicides were preceded by physical violence by the male partner before he killed the woman (Campbell et al., 2003). Intimate partner homicides of men by women are also characterized by a history of battering of the female homicide perpetrator by the male partner in as many as 75% of the cases (Hall-Smith et al., 1998; Campbell, 1992). It has long been noted that female-perpetrated intimate partner homicides are often characterized by self defense, when the male partner is the first to show a weapon or strike a blow and is subsequently killed by his victim (Block '93; Browne, Williams & Dutton, 1999; Campbell, 1992; Crawford & Gartner '92; Jurik & Winn '90; Smith et al., 1998; Websdale, 1999; Wolfgang, 1958).

Stalking

Stalking may be an even more common precursor of intimate partner homicide than abuse. McFarlane, Campbell and associates (1999) reported that stalking and harassment occurred in 70-90% of 200 actual and attempted femicides in 11 US cities. The strongest association was the combination of estrangement and prior abuse. Yet stalking also occurred in the majority of femicides in intact marriages and relationships where there was no history of violence.

Estrangement

An association has also been found between intimate partner homicide involving husbands and wives and a history of estrangement (Wilson et al., 1993, 1995; Dawson et al., 1998; Websdale, 1999). The estrangement may take the form of physical leaving or starting legal separation procedures. In an analysis of spousal homicide data from Canada, Chicago and the UK, Wilson and Daly (1995) found that the combination of physical and legal separation posed the most risk. From these studies and clinical experience with battered women, it has been theorized that male partners are threatened by loss of control over the relationship when women announce their decision to separate, and some men will stop at nothing to regain control, including femicide. Although it is clear that the period after separation is a time of increased risk, this danger has not been compared with the risk of staying in an abusive relationship. It has also been difficult to calculate whether or not separation increases risk in unmarried couples, since the proportion of separated to intact couples is not known. In studies in North Carolina and Ontario two to three times as many couples (married and not) were intact as were estranged when the woman was killed (Dawson & Gartner, 1998; Smith et al., 1998). The same proportion (70% intact, 30% estranged) was found in a recent preliminary data analysis of 100 femicides from the US 11-city study. However, in 20% of the "intact" couples, the woman had left and returned at least once in the prior year (McFarlane et al., 1999). If one considers separation at least once during the prior year as a more precise measure of estrangement, then 55% of the intimate partner femicide victims in the final study were estranged from their partners when killed (Campbell et al., 2003). In the final model, estrangement increased the risk of femicide (in comparison to abused women) by

an odds ration of 3.64. When the perpetrator was highly controlling *and* there was separation, the OR increased to 5.52.

Demographic Characteristics Associated with Intimate Partner Homicides

Like perpetrators of other homicides, male perpetrators of intimate partner homicides in the United States are mostly poor, young, a member of an ethnic minority group, have a history of other violence, and have a history of substance abuse (Weiner, Zahn, & Sagi, 1990). Resource theory of domestic abuse from sociology may offer a partial explanation. This theory suggests that when a man's personal resources, such as education, income, job prestige and community standing, are lower than his spouse's, he may use violence to decrease the perceived status difference (Walker, 1984; Howard, 1986). Often young ethnic minority males are poorly educated, unemployed, or underemployed in comparison with their female partners (National Institute of Justice, 1997; Jaynes et al., 1989; Bowman, 1993; Smith & Brewer, 1990). As a result, a small percentage may resort to violence and eventually murder as a means of exerting power and control to elevate or equalize their status in their intimate relationships. In the 11-city intimate partner femicide study, unemployment was the only demographic characteristic that was significant in the final model, increasing the risk of intimate partner femicide by OR = 4.42.

Guns

Retrospective and case control studies have associated the use of guns and substance abuse (both drugs and alcohol) with intimate partner homicide (Browne, Williams & Dutton, 1998; Campbell, 1995). Access to and availability of firearms in the US greatly increases the risk of homicide in general, as well as the risk of intimate partner homicide (Kellerman et al., 1993). Bailey and associates (1997) reanalyzed the results of two population-based case-control studies and found that prior intimate partner violence and gun ownership were strongly associated with femicide in the home. In a study of 134 homicides of American Indian, Hispanic, and Non-Hispanic white women in New Mexico, researchers also found that firearms were nearly two times more likely to be used in "domestic" (intimate partner) femicides than other femicides (Arbuckle et al., 1996). In most cities, handguns are the weapon of choice for intimate partner homicides (Wilt, Illman & Field, 1995), although in Chicago (Block et al., 1995), knives (37%) were the most commonly used weapon from 1993-96, with firearms a close second (37%). Male perpetrators were more likely than female perpetrators to beat an intimate partner to death and slightly more likely to use a handgun. In the 11-city femicide study, perpetrator access to a gun increased the risk of femicide by OR = 5.38 and then use of a gun drastically increased the risk of the worst incident of abuse being fatal (incident level risk factor) by OR = 41.38.

Alcohol and Drug Use

It is generally difficult to identify substance abuse in homicide perpetrators with certainty unless they have committed suicide and blood alcohol levels are thus available. There may also be differential risk depending on the substance used (drugs versus alcohol), general substance abuse versus intoxication at the time of the homicide, whether the victim or perpetrator or both were substance abusers, and sex of the victim and perpetrator. Males are more likely to be alcohol abusers as victims than as perpetrators, but male victims and perpetrators are more likely to abuse alcohol than are females in either category (Block, 1998; Campbell, 1992; Smith et al., 1998). Furthermore, alcohol use is not uniquely associated with intimate partner homicide for women. Both Wilt (1994) and Moracco (1998) found alcohol use in as many women killed by intimate partners as killed by others. Persuasive evidence about drug abuse was found in two large data

sets (Chicago and North Carolina) where significantly *less* drug abuse was found in cases of intimate partner homicide than other homicides (Moracco et al., 1998; Block & Christakos, 1995). Although both were significant at the bivariate level, neither perpetrator problem drinking nor perpetrator illicit drug use were significant predictors of femicide in the final models of the 11-city femicide study (Campbell et al., 2003). Illicit drug use was a stronger predictor than problem drinking and remained a significant predictor until perpetrator aggressive behavior toward the partner was added into the model. Although subsumed by more powerful predictors, a remarkable 70% of the male perpetrators were using drugs or alcohol at the time of the homicidal incident (Sharps et al., 2003). Neither victim drug nor alcohol use before or during the femicide were significant predictors after controlling for demographics (Sharps et al., 2001).

Mental illness

Zawitz (1994) reported that 13% of perpetrators (11% of males, 15% of females) in 540 intimate partner homicides in the US had a history of mental illness, compared to 3% (not reported by gender) of non-family murderers. In other data, approximately one-third of the 200 perpetrators in the 11-city study of attempted and actual femicides were described as being in poor (versus fair, good or excellent) mental health (Sharps et al., 2001b). Forty-six percent of the male perpetrators had had at least one contact with a mental health professional, as compared to 29% of the victims; 33% had had some contact with an alcohol or drug treatment program, as compared to 25% of victims. In the comparative analysis, perpetrator mental health was a significant predictor on the bivariate level but not in multivariate analysis. Few other studies have found mental illness to be a significant risk factor for intimate partner homicides, but few have been able to use adequate operationalization. There is more convergence of evidence of a history of mental illness of perpetrators in the studies of homicide-suicide.

Differential Risk by Gender: Homicide-Suicide, Overkill, Forced Sex & Abuse during Pregnancy

Homicide-suicides represent a significant proportion, 27-32%, of intimate partner femicides. This pattern is almost never seen when women kill a male intimate, only for only .1% of such cases in North Carolina (Block & Christakos, 1995; Moracco et al., 1998). Intimate-partner homicide-suicides follow different patterns than other intimate partner femicides. In this type of femicide, whites are disproportionately represented (Stack, 1997; Buteau et al., 1993; Websdale, 1999). Buteau and associates (1993) in Canada as well as US researchers (e.g. Morton et al., 1998; Stack, 1997) have divided homicide-suicides into "mercy killings" or "suicide pacts" that involve older couples afflicted by physical illness or other serious problems, on the one hand, and the more common case of intimate partner involuntary homicide followed by suicide of the perpetrator, on the other. In the much larger category of involuntary homicides involving younger perpetrators and victims (accounting for 90% of such cases in North Carolina), risk factors for perpetrators included being male, jealousy, current or past depression, a long standing relationship with the victim, a history of physical abuse or separation/reunion episodes, personality disorder, and alcohol abuse (Buteau et al., 1993; Morton et al., 1998). Separation was a factor in 45% of the North Carolina homicide-suicides (Morton et al., 1998). A significant proportion of the male perpetrators (15% in NC; 21% in Quebec) in the two major studies of the phenomenon had consulted mental health services in the year prior to the event (Buteau et al., 1993; Morton et al., 1998). Depression was reported in 46% and substance abuse in 23% of perpetrators in Quebec. In North Carolina, 38% of the homicide-suicide perpetrators had ingested alcohol before death, but this was a slightly lower

percentage than for partner femicides without suicide (Moracco et al., 1998; Morton et al., 1998). In the analysis of the 32% of femicides that were homicide-suicides in the 11-city femicide study, perpetrator threats of suicide and perpetrator history of poor mental health were unique predictors of this form of femicide (Koziol-McLain et al., under review).

Overkill is another characteristic of intimate partner femicide that is not usually present when a female kills a male partner (Campbell, 1992; Browne et al., 1998). Overkill was first described by Wolfgang in 1958 as two or more acts of shooting or stabbing or beating the victim to death. Several North American studies have found that the majority (46-90%) of women in intimate partner homicides are the victims of overkill, compared to 12% or less of males (Browne et al., 1998).

Descriptive evidence and preliminary case control analysis from the 11-city study suggests associations of forced sex and abuse during pregnancy with intimate partner femicide (Campbell, 1986; Campbell, Soeken, McFarlane & Parker, 1998; Campbell & Soeken, 1999; McFarlane, Soeken, Campbell, Parker, Reel, & Silva, 2002). Forced sex was a significant risk factor in the final model of the 11-city femicide study (OR = 1.8), but abuse during pregnancy was significant on a bivariate level only. Particularly violent and dangerous men may be those who also force their partners into sex and beat them during pregnancy. In addition, jealous and controlling men may suspect or have evidence that the unborn child is not their biological progeny and therefore may kill their partners out of male sexual competitiveness (Daly, Wiseman, & Wilson, 1997). This theoretical approach was supported by evidence linking heightened risk of intimate partner femicide (uxoricide) and the presence of stepchildren (Daly, Wiseman & Wilson, 1997). The importance of this risk factor was also found in the 11-city intimate partner femicide study with an increased risk of OR=2.4 (Campbell et al., 2003). However, the evidence can also be explained through a power and control framework.

Risk factors for Intimate Partner Violence and for IPRe-assault

Risk of Intimate Partner Violence

Early research on domestic violence focused on the risk factors for an individual to become an abuser, and there are several reviews of this body of knowledge. In this body of research, characteristics consistently found to be associated with violence toward an intimate partner include experiencing violence in the family of origin (either as a child victim of abuse or as witness to abuse of the mother by a partner), substance abuse, and unemployment (Gondolf, 1997; Gelles & Straus, 1990; Hotaling & Sugarman, 1990; Saunders, 1995; Tolman & Bennett, 1990). A number of other contributing factors have been posited but support has been inconsistent across studies. Although some studies find a strong relationship between these factors and battering, in other studies they do not reach significance or were not measured. These factors are stress, depression, low self-esteem, traditional sex role attitudes, personality disorders, and status inconsistency (the male partner having a lower status in regard to education, income, and occupation than the female partner).

Risk of Re-Offending

More recently, a separate area of inquiry into the risk of re-assault has begun to develop. Generally, it is only a minority of batterers who re-abuse physically during follow-up (Maxwell, Garner, & Fagan, 2001). Maxwell and colleagues (as well as Gondolf, 2000) conclude that there is

a small number of batterers that chronically revictimize their partners and that there is a need for research "to accurately predict high-rate repeat offenders and to find methods of helping their victims" (Maxwell, Garner, & Fagan, 2001, p. 36). Age is strongly associated with re-offending rates: that is, it appears that many batterers "age out" of battering their partners (Edleson, 1996; Maxwell, Garner & Fagan, 2001).

Most of the research that provides information about factors associated with repeat domestic violence have been studies of the impact of batterer programs on re-offending. The strongest of these studies have compared offenders mandated to a batterer program to offenders not so mandated, in an experimental design. Generally, these studies have found re-offending rates of physical abuse based on victim reports in the range of 30% to 50% and from 10% to 40% based on criminal justice data Percentages vary depending on the length of the follow-up period and measures used. A recent experimental study in Broward County, Florida, found a high re-arrest rate of 24%. The strongest predictor of re-arrest was a lack of "stake in conformity": that is, whether the perpetrator was consistently employed, was married and had a stable residence (Jackson, Feder, Forde, Davis et al., 2003). Fourteen percent of the victims reported severe battery during the one year follow-up. Again, a lack of stake in conformity best predicted severe physical abuse by victim report: "women involved with, but not married to, younger jobless men were more likely to report incidents of severe physical violence." Previous arrest and incarceration were another of the key indicators of reassault (Jackson, et al., 2003). Gondolf's (1997) multi-site study of men in batterer programs implicated alcohol, "voluntary" (as opposed to court-mandated) program participation, previous arrests, and severe psychological problems as risk factors for reassault. Saunders (1995) identified violence outside of the home and lack of remorse as indicators of risk of severe abuse. Dutton (1997) found that men who completed a batterer program were more likely to re-assault their partners if they had antisocial, borderline and avoidant personalities than if they did not score high on these traits.

Attempts have been made to create typologies of abusers, with different types posing different levels of risk. Holtzworth-Munroe and Stuart (1994) have synthesized this body of research and proposed and then tested (Holtzworth-Munroe et al, 2000) a typology with 3 categories of abusers: the antisocial, the "ordinary" batterer, and the dysphoric-borderline type. According to this typology, the antisocial batterer (somewhat analogous to Jacobson's [1996] Type I or "cobra" type) would be at high risk for committing repeated assaults and perhaps homicide; the dysphoric-borderline batterer would be at risk to commit homicide and perhaps homicide-suicide if the partner left but is otherwise less assaultive and less recognizable as dangerous; and the "ordinary" batterer would be at lower risk for re-assault or homicide.

Victims' Predictions

Recent studies have found that women's perception of risk is important in determining risk of re-assault by an intimate partner. Weisz, Tolman, and Saunders (2000) found that women's perception of danger was the single best predictor of re-assault, a stronger predictor than any of the 10 items from the Danger Assessment available in criminal justice records. Similarly, Goodman, Dutton and Bennett (2001) in a sample of 92 women found that women's prediction of re-assault was the strongest single predictor of re-assault. In an analysis of their data on 499 men in batterer intervention programs and their partners, Heckert and Gondolf (2004) found women's perception of risk to be a significant predictor of revictimization by an intimate partner, stronger than the SARA (Spousal Assault Risk Assessment) and K-SID (Kingston Screening Inventory for

Domestic Violence) but not as strong as a simulated version of the Danger Assessment. The best model of prediction was the DA along with women's perception of risk. However, in the 11-city femicide study, only 47% of the actual femicide victims and 54% of the victims of attempted femicide accurately assessed that their perpetrator was capable of killing them.

Summary of Risk Factors and Implications for Risk Assessment

Risk for Intimate Partner Homicide vs. Re-Assault

Clearly, the patterns of intimate partner homicide vary systematically according to the sex of the perpetrator. Also clear is that the current body of research supports some risk factors, although specific individual and household risk factors are as yet not fully determined – even in North America where the majority of research has been conducted (Kellerman et al., 1993; Campbell, 1995; Browne et al., 1998; Campbell et al., 2003). That is, some factors seem to be generally present among men who batter, and there seem to be differentiating characteristics among those whose assaults escalate, and those who commit homicide. At the same time, other factors are not consistently supported by the research, and some of the key factors are present in the background of men who do not batter.

Prior domestic violence and handguns are the most consistent and strongest risk factors for homicide, with estrangement, a stepchild in the home and unemployment also strongly implicated. Although violence outside of the home, alcohol abuse, and minority ethnicity are also implicated in male perpetrated intimate partner homicide, they seem to be less strong risk factors than for other types of homicide. Female perpetrators are far less likely to have had a history of perpetrating any kind of violence. Homicide-suicide with a male perpetrator is a form of homicide especially associated with intimate partnership. Guns, estrangement, and prior mental health problems in the form of depression or suicidality are particular risk factors for this form of intimate partner homicide. Other aspects of the intimate partner relationship, such as abuse during pregnancy and stalking, have also been implicated as risk factors, but the only national comparison study to date (Campbell et al., 2003) did not find them to be among the factors in the final models.

Although there is overlap between the risk factors for re-assault by an intimate partner and the risk factors for intimate partner homicide, there seems to be a difference of degree and some differential patterns. For instance, substance abuse is more of a risk factor in domestic assault and re-assault than in domestic homicide, while perpetrator suicidality is more of a risk factor in murder of intimate partners by men (because of the large proportion of homicide-suicides) than in murder of intimate partners by women or in domestic violence re-offending. Child abuse victimization and witnessing domestic violence in childhood are well documented as risk factors for intimate partner abuse perpetration and therefore are presumed to be risk factors for re-assault (Straus & Gelles, 1990; Hilton, Harris, Rice, Lang et. al., 2004). However, neither has been implicated in intimate partner lethality, perhaps because this history generally is not part of homicide records. Gun ownership, clearly implicated in domestic homicide, does not seem to be a risk factor for either minor or severe wife assault (Hanson et al., 1997).

Clinical vs. Actuarial Assessment of Intimate Partner Violence

Since a history of domestic violence against the female partner is the most consistent and strongest risk factors for intimate partner homicide, one of the major approaches to decreasing this

form of mortality is to assess risk with battered women and male perpetrators of domestic violence. Clinical assessment of risk has been recommended and used by domestic violence victim advocates for more than a decade (e.g., Hart, 1988). Many professionals, including those in law enforcement, victim services and health care are now looking for actuarial methods for risk assessment; that is, instrumentation based on empirical item selection and with fixed, explicit and validated formulas for assignment of risk categories and decision making (Grove & Meehl, 1996; Hilton, Harris, Rice, Lang et al., 2004). As described by Hilton and colleagues (Hilton et al., 2004) these formal risk assessment methods are different from the usual psychological tests in that the usual methods of instrument item selection and psychometric evaluation (factor analysis, item analysis, internal consistency, test-retest reliability, construct validity) are less relevant. Instead, empirical item selection (using multivariate analysis from an existing data base), assessing interrater reliability, cross-validation or discriminant group validity and most importantly predictive validity are the most salient evaluation techniques.

Because of the low base rate of intimate partner homicide and the ethical necessity to intervene in obviously dangerous cases, we will never be able to predict intimate partner homicide or determine predictive validity of any instrument purporting to assess lethal risk with total certainty (Campbell, 1995; Litwack & Schlesinger, 1998; Mulvey & Lidz, 1995). In the general field of violence, there is recognition that risk assessment is the appropriate terminology for the enterprise, rather than prediction of violence or prediction of homicide (Monahan, 1999; Hare, 1999). There is also recognition of the importance of improving our ability to determine risk, however difficult and uncertain, because of the seriousness of the potential outcomes -- loss of human life or loss of liberty -- (Mulvey & Lidz, 1995).

The first instruments specifically assessing risk in the domestic violence field were developed to assist battered women and advocates with safety planning, and therefore were not intended to predict risk accurately but more to create awareness in the victim and elicit information for the advocate. Subsequently, a number of instruments were developed for use in the criminal justice system to evaluate the risk of recidivism, modeled on other instruments for other categories of criminals. Most recently, more in line with the instruments designed to assess risk of sexual assault recidivism and community violence by the mentally ill, some instruments have been designed to assess risk of intimate homicide specifically. The latter are usually designed as actuarial methods but do not yet have sufficient psychometric data support to be used as the sole determinant of decision making.

Review of Intimate Partner Violence Risk Assessment Instruments and Assessment Methods

Roehl and Guertin (2000) compiled an overview of instruments designed to assess dangerousness in domestic violence situations and their use in sentencing. Ten instruments and one threat assessment method were identified and reviewed: the Danger Assessment (Campbell, 1992; 1995); Domestic Violence Inventory; Domestic Violence Risk Assessment Form; Domestic Violence Screening Instrument (Williams, 1999); Kingston Screening Instrument for Domestic Violence (Gelles, 1998); MOSAIC-20 (De Becker, 1997); Lethality Checklist and Physical Abuse Scale; Pre-Sentence Investigation (Pence, 1999) Domestic Violence Supplement; Risk Assessment and Lethality Assessment; and the Spousal Assault Risk Assessment (Kropp et al., 1995; Kropp & Hart, 2000). In addition to those reviewed by Roehl, the Navy Risk and Safety Assessment was also designed to be an actuarial instrument but has not yet been evaluated.

Subsequently, the O.D.A.R.A. (Hilton et al., 2004) was developed with claims to be the only true actuarial DV risk assessment instrument since it was developed based on criminal justice records (in Ontario). It was validated prospectively with an ROC curve assessment of .77. Validity data on the PAS (Propensity for Abusiveness Scale) designed as the only risk assessment instrument to address potential cases before they occur has also been published since the Roehl and Guertin overview (Dutton, Landolt, Starzomski & Bodnarchuk, 2001), and there has been an additional review of these instruments (Dutton & Kropp, 2000).

Two additional "clinical" domestic violence risk factor lists are examples of those widely used in the field in various settings that were developed as clinical instruments rather than with psychometric evaluation in mind. They have often been adapted in various ways and, as far as is known, are not involved in any ongoing studies. Hart's (1988) domestic homicide risk factor list has been widely used as the basis for safety planning in shelters and other victim service settings, while Sonkin's (1985) instrument was the first published and developed for use in batterer treatment settings.

An important distinction can be drawn between assessing risk of re-offending vs. risk of lethal violence. This distinction influences which risk factors are identified, intended use of the instrument/method, and its validation method. The Danger Assessment (DA) and MOSAIC-20 were specifically designed to assess risk of lethal or near lethal violence, while most of the other instruments, such as the Kingston Screening Instrument for Domestic Violence (K-SID), Domestic Violence Screening Inventory (DVSI), SARA and Presentencing Inventory (PSI) were designed to identify risk of re-offending in spousal assault cases. (Three other instruments – Domestic Violence Risk Assessment Form, Lethality Checklist and Physical Abuse Scale, and Risk Assessment and Lethality Assessment – have separate lethality risk factors and re-assault or abuse-related factors.)

Other differences follow from the purpose and intended use of the two types of approaches. The lethality assessments were designed to be used primarily for prevention, with information from victims. In contrast, most of the instruments designed to assess risk of re-assault were developed for use in the criminal justice (or military) system for sentencing, probation, bail and treatment decisions, with the offender as the primary respondent. For lethal risk assessment, ease of administration and brevity are extremely important. In criminal justice settings, more actuarial evaluations are feasible and important to conduct and many sources of data can be obtained. Both types of approaches are important and may involve different factors. It is far more difficult to assess predictive validity of the lethality assessment systems, however, because of the extremely low base rate of domestic homicide.

Description and Validation of Instruments and Threat Assessment Methods

Published data on the instruments is meager and independent evaluations are practically nonexistent (Roehl & Guertin, 1998, 2000). Many of the instruments/threat assessment methods are based on risk factor lists with no empirical research, but the MOSAIC-20 (De Becker, 1997) and the DA (Campbell, 1992; Campbell, 1995) were based on homicide case evaluations and research. The Domestic Violence Risk Assessment Form and the K-SID (Gelles, 1998) were based on independent domestic violence case evaluations. The development of the K-SID also utlized survey research. According to Roehl and Guertin (2000) and personal communication with authors, the SARA (Kropp, Hart, Webster, & Eaves, 1995; Kropp & Hart, 2000), K-SID (Gelles,

1998), DVSI (Williams, 1999), PSI (Pence, 1999) and DA (Campbell, 1992; Campbell, 1995) are currently part of ongoing validation studies with some preliminary data having been presented at conferences (Gelles, 1998; Kropp, 1998; Kropp & Hart, 2000; Williams, 1999). Validity data on the DVSI were recently published (Williams & Houghton, 2004). A study of the Navy Risk Assessment is also underway. Psychometric data on the DA (Campbell, 1995) have been published, and some independent research using the instrument has been conducted which will be reviewed below. Since this study was designed, description and data on two instruments have been published. The O.D.A.R.A. (Ontario Domestic Abuse Risk Assessment) was developed through a large Canadian database of domestic violence offenders and promising predictive validity data have been published (Hilton et al., 2004). The DVI and SARA are being used fairly widely but will not be tested in the proposed experiment. They will be described briefly below.

SARA

The SARA (as well as the DVI) was designed to assess risk of re-offending in the criminal justice system. The SARA is described as an instrumental means of coding professional judgment (Kropp & Hart, 2000). A complete evaluation with the SARA requires psychological assessment of the perpetrator and clinical judgment. The SARA is currently being used widely in Canada and Vermont for probation supervision and treatment decisions and as part of a validation study in Colorado (Kropp & Hart, 2000; Roehl & Guertin, 2000; Williams, 1999). It was developed in the early 1990's and is described as an assessment guide based on a thorough literature review, rather than as an actuarial instrument or psychological test (Kropp et al., 1995). It is recommended for use as part of an in-depth assessment to be used for judicial and probationary decisions. It has four different scoring approaches, including a 0-2 assessment of as well as absolute presence of each of 20 risk factors (with space for the evaluator to add other "considerations" which can also be rated or counted). In addition, there is a score consisting of the evaluator's judgment of how many of the risk factors (including the additional considerations) are "critical" and an independent overall summary rating of risk to the partner and to others. The manual urges the evaluator to use as many sources of information as possible in filling out the SARA, including both victim and perpetrator interviews, additional criminal justice records, and standardized instruments, especially the Hare Psychopathy Checklist (Hare, 1991).

Data on the SARA is based primarily on testing in Canada, with a primarily (80%) white population. In six samples (total N = 2681), the SARA has shown good internal consistency (Alpha = .62-.83) and item homogeneity (Kropp & Hart, 2000). Discriminant validity was strong: in a large sample (N = 1,010) of inmates, the SARA discriminated well between those with a spousal assault history and those without. In a smaller study (N = 102) of recidivists versus non recidivist spouse assaulters, the number of critical items score discriminated best between the two groups. In tests of construct validity, correlations of the spousal assault risk with other instruments assessing general risk of violence were moderate to low. Part of the explanation for these findings is that many spouse assaulters are not necessarily generally violent (Holtzworth-Munroe, & Stuart, 1994).

In a more recent prospective study of 1465 male domestic violence offenders in Colorado (Williams & Guerra, 2003), the SARA was significantly associated with risk of reoffending according to reports from 125 female partners after six months and review of criminal justice records after 18 months. The SARA correctly identified 66% of the cases for overall recidivism.

The Domestic Violence Inventory – Risk and Needs Assessment (DVI)

The DVI is designed not only to assess risk of re-assault but also to assess treatment needs. Unlike the SARA, the DVI is a structured questionnaire to be completed by the perpetrator. It takes 30-35 minutes to complete, has been translated into Spanish and information is available on the instrument on its web site. There are seven scales including one for truthfulness (the only such scale in the variety of instruments), alcohol and drug abuse scales, violence potential, and treatment needs. The scores are divided into categories of low, medium, problem, and severe risk. Self-published data are reported for more than 5000 respondents, but there have been no refereed publications on the instrument. Convergent construct validity shows correlations ranging from .54-.78 with the MMPI. Reliability is supported by an internal consistency of .84-.91, but inter-rater reliability is low, ranging between .10 and .54.

K-SID

The K-SID was developed from the extensive program of research of Richard Gelles (Gelles & Straus, 1990) as a screening instrument. It was explicitly designed as an actuarial assessment, although it does not yet have enough evaluation to be considered such. The K-SID consists of ten risk markers for re-assault with three categories of response for each item. Certain response categories merit points that are added for classification into low (0-3), moderate (4-6), high (7-10) and very high risk of re-offending. The summary score is not intended to be used for anything other than risk category classification and is not intended for use as a continuous variable. and didn't want the summary score to be used as a continuous variable. In the original instrument, the presence of either of two additional risk factors – previous domestic violence arrest and previous violation of a temporary or permanent protection order – put the perpetrator into the very high risk category regardless of the point total. After predictive validation testing by Lyon (1998), the previous domestic violence arrest item was eliminated as an automatic high-risk qualification. There is also a poverty chart to determine the answer to the income risk factor, a formula for determining “binge” alcohol use, and a severity and injury chart that does not factor into the score for risk classification.

Gelles (1998) has reported data from several ongoing studies indicating support for internal consistency and discriminant group validity. Lyon has tested predictive validity over two years on a large sample of arrestees. On the basis of police and court records for offenders being arraigned in three Connecticut courts in 1996 (n=1014), Lyon found that the K-SID score significantly predicted re-arrest within one year. In a two-year follow-up, Lyon found that a “high-risk” classification significantly predicted a new domestic violence arrest, and that the overall score predicted better than a history of family violence arrests. In the “moderate,” “high” and “very high” risk categories on the basis of total scores, 34-38% were re-arrested within two years ($p<.0001$). However, these three categories were very similar and Lyon concludes, therefore, that the instrument was most effective at predicting low risk: only 12% of those classified as low risk were arrested for a new family violence crime within two years. The K-SID is being used throughout Connecticut as a basis for criminal justice decisions including probation, incarceration and protection order conditions.

DVSI

The Domestic Violence Screening Inventory (DVSI) was developed by Kirk Williams and is being used by Probation in Colorado to determine level of probation supervision. It is intended for use as a screen to be followed by more intensive evaluation with the SARA if the DVSI score suggests a high level of risk. The DVSI consists of twelve questions primarily related to criminal history of

domestic violence and other violent offenses, including arrests, convictions and violation of restraining orders. Employment status, recent separation, and presence of children during the incident are the other risk factors probed. A predictive validity test on the basis of criminal history was recently completed in a prospective study of 1465 male domestic violence offenders in Colorado (Williams & Guerra, 2003). The DVSI was significantly associated with risk of reoffending according to criminal justice records after 18 months and reported violence after 6 months from the 125 female partners who provided information. Sixty-threepercent of the cases of overall recidivism were correctly classified (compared to 66% for the SARA, as above).

Danger Assessment The DA is a measure designed to assist battered women in the assessment of their own risk of femicide (Campbell, 1986). The Danger Assessment (DA) was originally developed by the PI of this study (Campbell, 1986) with consultation and content validity support from battered women, shelter workers, law enforcement officials, and other clinical experts on battering. The initial items on the Danger Assessment instrument were developed from retrospective research studies of intimate partner homicide or serious injury (Campbell, 1981; Browne, 1987; Berk et al., 1983; Fagan et al., 1983).

The first portion of the measure assesses severity and frequency of battering by presenting the woman with a calendar of the past year. The woman is asked to mark the approximate days when physically abusive incidents occurred, and to rank the severity of the incident on a scale of 1 to 5 (where 1=slap, pushing, no injuries and/or lasting pain and 5=use of weapon, wounds from weapon). The calendar portion was conceptualized as a way to raise the consciousness of the woman and reduce the denial and minimization of the abuse, especially since using a calendar increases accurate recall in other situations (Campbell, 1995; Ferraro et al., 1983). In the original scale development, 38% of women who initially reported no increase in severity and frequency changed their response to “yes” after filling out the calendar (Campbell, 1986; Campbell, 1995). The second part of the DA was a 15-item yes/no dichotomous response format of risk factors associated with intimate partner homicide. Both portions of the instrument take a total of approximately 20 minutes to complete. The woman can complete the instrument by herself with professionals from the health care, criminal justice, or victim advocate systems assisting in the interpretation of the instrument within the context of her situation. The original DA was scored by counting the “yes” responses with no classification or cutoff score; a higher number indicates that more of the risk factors for femicide are present in the relationship.

The initial studies using the DA instrument are described in detail elsewhere (Campbell, 1986, 1995). In the 15 research studies with reported data in which the DA has been used to date, reliability ranged from 0.60 to 0.86 (Campbell, 1986; Campbell, 1995; McFarlane et al., 1992; McFarlane et al., 1996; McFarlane et al., 1998; Woods, 1999; Silva et al., 1997). In the recent 11-city femicide study, internal consistency ranged from .74 in the sample of 496 abused controls, to .75 in the sample of 183 attempted femicide victims and .80 in the 263 actual femicide sample. In two studies in which test-retest reliability was assessed, it ranged from 0.89 to 0.94 (Stuart et al., 1989; Campbell, 1994). All samples include a substantial portion of minority women (primarily African-American and Latina) and women from a variety of settings. Internal consistency reliability estimates did not vary for African-American, Latina, and White women (Campbell, 1995).

Discriminant group construct validity was supported by significant differences in mean scores among contrasting groups of women. The lowest mean score was found in the non-abused sample

with the highest score among the sample of women in the emergency department (Campbell, 1995). In addition, the DA successfully discriminated between groups of abused and nonabused women in two studies (Woods, in press; McFarlane et al., 1998). Convergent construct validity has been supported in the majority of the studies with moderate to strong correlations with instruments measuring severity and frequency of domestic violence, including the Index of Spouse Abuse, the Conflict Tactics Scale and injury from abuse (Campbell, 1995; McFarlane et al., 1998), as well as Post-Traumatic Stress Disorder (PTSD) (Woods, in press; Silva et al., 1997).
The 11-city case control study (comparing actual and attempted femicide cases with those of physically abused or threatened women and non abused women) provided further validity data for the DA and indicated appropriate weighting of the items. The only factors from the DA that did *not* show robust significant differences ($p < .05$; % differences >20) were child abuse (using the narrow operationalization of reported child abuse), *victim* suicidality, and perpetrator violence outside of the home (37-39% in the latter two categories). A factor not on the DA that was a significant differentiating item was stalking and harassing behaviors. The original DA demonstrated significant differences in mean scores between the abused controls and the actual and attempted femicide groups.

A modification of the DA based on the results of 11-city femicide study with weighted scoring was submitted to ROC curve analysis with the area under curve .90 among the actual femicides and then tested among the attempted femicides resulting in a .92 area under the curve. Sensitivity (proportion of cases or women assaulted during follow-up who were correctly identified as high risk) is good (90%) at the lowest scoring level (named "variable danger") with 86%, 83% and 57% at the next three levels of danger ("increased," "severe" and "extreme danger") among the actual femicides (and similar percentages among the attempteds). Specificity (abused controls) was 69% at the lowest (variable) level, 70% at Increased Danger, 80% at Severe Danger and 98% at the Extreme level of danger. In other words, only 2% of the women abused but not killed scored in the extreme level of danger, 20% in the Severe Danger range; 30% at Increased Danger and 31% at the Variable Danger Range.

In addition, three independent predictive validity studies have been conducted of the DA. In one, Goodman and colleagues (2000) found that the DA successfully predicted reassault in a sample of 92 battered women who had filed a protective order. Although there was only a 53% retention rate of the sample, the DA was a stronger predictor than the Revised Conflict Tactics Scale (Straus, 1997), with a 4.2 vs. 2.8 odds ration per SD. Weisz, Tolman and Saunders (2000) found that women's perception of danger was a stronger predictor than any of the 10 DA items (of the 15 on the original scale) available in the criminal justice record. However, it needs to be noted that only 2/3 of the items on the DA were used and they were evaluated singly rather than in combination as intended. Even those items were not actually asked of the female partner as intended but coded from criminal justice records using slightly different operationalizations in several cases. In fact, the DA asks about victims their perception of perpetrator homicide capability, but this item was not included in the Weisz et al. evaluation. In an evaluation of 499 perpetrators in the criminal justice system, Heckert and Gondolf (2004) found that the items from the DA they were able to assess had a sensitivity of 66% but there were 33% false positives. Women's perception of risk plus the DA was the best model (over SARA & K-SID) for predicting reassault, but women's perception of risk by itself was not quite as predictive as the DA.

DV-MOSAIC

The DV-MOSAIC is a computer-assisted method of threat assessment developed by Gavin de Becker & Associates (de Becker, 2005). DV-MOSAIC is not a predictive instrument. It is an overall method used to aid police officers in their assessments and investigations of domestic violence situations. Only part of the MOSAIC method is designed to apply predictive strategies to police officer and victim decision making. DV-MOSAIC aids in the assessment and consideration of hundreds of individual features of a domestic violence situation. Some questions and features are included because they are required components of police investigations, and are not expected to be predictive. Other questions and features are included because the developers and Development Team advisors believe they will enhance the quality of police investigations and assessments, or because they are perceived to increase the safety of a victim of abuse, and not because they are expected to be predictive. Some areas of inquiry are included within MOSAIC because the developers perceived them to be predictive on the basis of existing research, expert opinion, practitioner opinion, and practitioner experience. It is the ability to assess the likelihood of reassault of these specific areas of inquiry within DV-MOSAIC that are being evaluated as part of this study.

As previously described, DV-MOSAIC was designed to be part of a comprehensive police investigation and assessment, rather than an instrument in the traditional sense. Answering the questions or probes that are part of the DV-MOSAIC threat assessment process is designed to assess "the likelihood of escalation or violence between the person being pursued and the person under assessment," according the introduction to MOSAIC manual, a confidential document. It is designed to evaluate the risk of escalation, continued violence and homicide by a man against his wife or girlfriend. De Becker emphasizes that the use of DV-MOSAIC is part of a process of risk assessment, a way of offering "the victim an opportunity to consider an organized overview of her situation." "The intent is to improve the quality of assessments…to contribute to the" assessor's "prediction" so that the assessor will then intervene effectively and also present the results to the victim so that she is better informed about the need for intervention. The introduction stresses that dangerousness is situational, not a permanent or stable attribute. DV-MOSAIC considers a violent disposition irrelevant, because we all have the potential to be dangerous, and people who lack a violent disposition may be dangerous under certain circumstances. Furthermore, for danger to be a present risk, there must be not only "sinister intent," but also the ability and means to carry out the intent, and access to the victim. Harm is not restricted to physical injury: causing anxiety and fear and invading privacy is also harmful. Dangerousness applies not to an individual who may have shown warning signs, but to a *relationship* between the individual and his target.

The manual warns that an accuracy rate with 10% false positives (people judged dangerous who will not, in fact, commit violence without any intervention) and 15% false negatives (people judged not dangerous but who will commit serious violence if there is no intervention) is unacceptable if applied to the population at large. The result would be to restrict the freedom of innocent people and to miss a number of cases of violence. The results would be no worse without risk assessment. De Becker argues, however, that the threat assessment method is not intended for use with the population at large, but only for assessing individuals who have already demonstrated an inclination toward violence against another specific individual. The MOSAIC is not used for intervention with the person (potential perpetrator) being assessed beyond those otherwise justified by criminal behavior. Robert J. Martin, Vice President of Gavin de Becker & Associates, and a veteran of the LAPD, describes the MOSAIC as a case screening assessment, a filter to accompany other forms of investigation.

The DV-MOSAIC consists of 46 questions (an additional question asks whether the victim will see the report) with multiple response categories. The response categories are sometimes complex, covering constellations or patterns rather than single features of the perpetrator or relationship. The program produces two scores: an IQ (Information Quotient) score, ranging from 1-200, based on the amount of information available for the assessment, and a Rating from 1-10, with 1 representing the lowest level of risk and 10 the highest. Both scores are based on item weighting. Martin gives the example that the age of the perpetrator might be worth 2%, but knowing about the perpetrator's ownership of weapons might be worth 6%: the "weapons" item therefore contributes more value to the IQ score. According to the developers, the 1-to-10 scale was suggested by MOSAIC co-developers at the U.S. Marshals Service because it is clear, familiar, and has consistent meaning to various users. They go on to state that it is not meant to imply that a 7 is some precise scientific amount less significant than an 8 -- in other words, it is meant to be an ordinal, not ratio level scale. They also state that it is not meant to prescribe that some specific case-management response should follow a specific rating. Threat assessment practitioners are meant to use all their information and insight when making case-management decisions. The MOSAIC threat assessment purposefully uses the numerical scale rather than words to describe levels of threats to avoid the connotations and conclusions that words can convey, as opposed to allowing assessment practitioners to express conclusions based on more than just rating.

The DV-MOSAIC measures many but not all the factors associated with risk of escalation or pursuit. The program is described as using a case-matching or template-matching procedure to produce the rating of risk. That is, the combination of features "matches cases that involved escalation and pursuit of encounters" or is "most similar to cases that have not escalated or pursued encounters" (Martin, MOSAIC-20 Manual, p. 2 of MOSAIC Overview). It assigns values to interrelated factors and compares them to cases where the outcome is known.

The DV-MOSAIC produces a report for each assessment, giving the IQ score and rating, and a sentence describing the similarity of the profile to cases that have escalated to homicide (for a high score) or have not escalated to homicide (for a low score). The report also describes the answer to each item and explains the relevance of that item to assessing risk of homicide with the intent to provide education and training.

The development of MOSAIC is described as involving study of over 250,000 "communications" and 18,000 cases. It is not clear whether this description applies to all the MOSAIC programs – there are other programs for assessing threats against government officials, workplace violence, school violence and stalking – or specifically to DV-MOSAIC, designed for estimating risk of violence against women by husbands or boyfriends. The DV-MOSAIC was based on cases of intimate partner homicide from the LA police department. A utility of MOSAIC, according to its developers, is that it provides uniformity to assessment such that ten different people of different abilities and styles will come up with the same preliminary rating. However, no inter-rater reliability data are provided. In regard to accuracy, the developers state that a perfect predictive performance cannot be expected, but case management decisions can be improved by use of the system. The MOSAIC approach is similar to the approach Kropp (2004) endorses and describes as a "structured professional judgment." De Becker and Associates also stress that the issue in threat assessment is prevention, not prediction, and that practitioners should seek to discover and

recognize the factors associated with escalation - and then intervene or encourage effective interventions in order to influence outcomes.

The three instruments and one threat assessment method are summarized in Table B1.

Table B1: Description of Four Risk Assessment Methods

Method	Description	Administration	Primary intended uses
Danger Assessment (Campbell, 1986, 1995, 2003)	Review of past year with a calendar[1] to document severity and frequency of battering and 20 yes/no questions about risk factors. Scoring: -3 -40 & four risk categories (variable, increased, severe & extreme danger).	Interview with the victim, usually by victim advocate.	Assess risk of extreme dangerousness and lethal violence for victim education, awareness, safety planning and service provision.
DV-MOSAIC (Gavin de Becker & Associates, 2001)	Computer-assisted method that includes 46 multiple response items about risk and protective factors. Scoring: Program computes risk score of 1-10 and a missing data (IQ) score.	Criminal justice professional enters responses after victim, perhaps offender and other interviews; reviews of criminal records and police reports. [2]	Assess immediate, short-term threat of severe or lethal domestic violence situations for victim awareness, safety planning, further investigation, and criminal justice responses.
Domestic Violence Screening Instrument (Williams & Houghton, 2004)	12 questions given 0 - 3 points, primarily related to offender's criminal history, employment, & several other risk factors. Scoring: Risk score 0-30, and two risk categories (not high risk & high risk).	Probation or other court officer completes instrument based on offender's criminal record and interview.	Assess risk of recidivism/re-assault for supervision, probation/parole, and other offender-related decisions.
Kingston Screening Instrument for Domestic Violence (Gelles, 1988)	10 questions about risk factors, each with 2 to 3 response categories, and a offender's poverty status scale. Scoring: Risk scores of 0 to 10 and four risk categories (low, moderate, high, or very high).	Offender and victim interviews and review of police reports by probation or other court officer.	Assess risk of recidivism/re-assault for offender charging and supervision decisions, set conditions for release, probation, and protective orders.
Victim assessment of risk (Goodman, Dutton 2000; Heckert & Gondolf, 2004; Weisz, Tolman, Saunders 2000)	2 questions about victim's perception of the likelihood that she will be physically assaulted or seriously hurt by abuser in the next year Scoring: victim rates likelihood on a scale of 1-10		

[1]The calendar portion of the DA was not used in this study.

[2] The DV-MOSAIC "domains of inquiry" were reformatted by the investigators as a victim interview

Need for Assessment of Risk in Intimate Partner Violence

Benefits to the Criminal Justice System

Much progress has been made in developing risk assessment instruments and systems in the field of intimate partner violence. Specifying the type of violence increases the accuracy of prediction, and there is no doubt that instruments specific to this field are needed (Monahan, 1984). For instance, psychopathy (Hare, 1991) has been found to be strongly predictive of much of violent behavior, but Kropp has not found it to be even moderately related to the risk of intimate partner reassault using the SARA (Kropp & Hart, 2000). Another strategy to increase statistical accuracy

of instruments is to increase the base rates of the behavior being predicted by using re-assault rates, because about 25-50% of batterers re-offend, as compared to 3% who commit potentially lethal violence (Dutton, 1995) and 2). Measurement sensitivity can also be enhanced by going beyond arrest as an operationalization of reassault and, for example, using partner reports of violence.

In order to improve the assessment of risk in intimate partner violence, more testing is needed of the instruments and methods already developed. Great ingenuity is needed in order to design evaluations that still provide for the safety of victims. Instruments also need to be evaluated in terms of their cultural appropriateness. Going beyond ordinary translation of instruments, attention needs to be paid to whether or not the actual risk factors make sense in different populations and whether psychometric properties hold in different ethnic and linguistic groups.

One Specific Potential Victim – Safety and Process Issues

For much of violence prediction, the safety of the general public is at risk rather than that of a specific individual. In contrast, intimate partner violence risk assessment relates to risk to one particular victim (or as in the case of a change in partners, a second person). It has long been recognized that the battered woman victim has to be central to the process of intimate partner risk assessment, in terms of both her perception of risk and safety planning (Campbell, 1986; Campbell & Stuart, 1989; Hart, 1988). Many batterer programs are in touch with the partners for purposes of safety planning and service referral. Some programs also obtain information from partners about reassault and psychological abuse either for program evaluation or for monitoring of progress. There is concern about victim safety, however, when victims are used to inform on progress or recidivism to the program, especially when new offenses must be reported back to the court or probation (O'Sullivan, 1998).

All of the systems that interact with battered women – health care, criminal justice, and social service – need to keep in mind the "duty to protect" as well as duties to warn (Campbell, 1994; Hart, 1988). All of the systems have responsibilities in terms of decisions about what is done with the perpetrator but also what protections are in place for victims. Risk assessment has the potential to serve a critical role for law enforcement, the judiciary, probation, victim advocates inprosecutor's programs, and the rest of the criminal justice system, as well as shelters, other victim advocacy organizations, and ED and other health care professionals in aiding decisions about what actions they need to take. Professionals, agencies and communities that have adopted or are planning to adopt formal methods of risk assessment need to know which instruments and approaches have statistical support for their accuracy and which are most useful in terms of ease of implementation for their particular purpose. However, no method of prediction will ever be perfect. Victims must all have basic safety measures available to them, access to police when a crime is committed, and safe shelter, no matter what a risk assessment indicates.

Conclusions

Continued efforts to develop a valid and comprehensive list of risk factors for intimate partner homicide are necessary. The field is seeking more accurate assessment of potential lethality in domestic violence relationships than the instruments and systems available can provide. Clinicians have been held accountable for not adequately predicting and protecting clients from potential danger (Hart, 1988; Litwack & Schlesinger, 1999; Mulvey & Lidz, 1995). In medical settings, legal experts generally agree that if a patient is a serious threat to someone else, the therapist or health

care provider must warn the potential victim (Campbell, 1994). Many domestic violence experts also feel a responsibility to conduct a separate assessment of the potential risk for homicide for victims as a basis for safety planning (Hart, 1988). Use of an instrument or threat assessment method may augment clinical or professional judgment in all the relevant systems (criminal justice, victim advocacy and physical and mental health systems) by ensuring that the risk factors specific to intimate partner homicide are assessed. In addition, an effective "coordinated community response" of victim assistance, batterer programs, community agencies, courts, police, probation, and health care could be improved by having a common language to indicate elevated risk and a way of selecting cases for intensive responses.

Actuarial methods for risk assessment in intimate partner violence situations are beginning to be tested but this testing is only in the beginning stages. Well-substantiated instrumentation, with independent validation and the kind of meta-analyses available in the area of sexual assault are still several years away. Several instruments have been developed, but as yet published psychometric data are sparse and predictive validity studies results are only beginning to be published on any of the instruments. The research reported here significantly advances the field by providing a prospective validity evaluation of three of the instruments and the one threat assessment method using sophisticated analysis techniques. This advance will be important in addressing the safety of battered women.

C. METHODS

The methodology of the study was largely developed during the planning study. They were complicated, involving the extensive development of measures, arrangements with multiple sites, hiring and training of staff, planning recruitment of participants, etc. At all times, one of the paramount goals was to keep victims safe.

Measures

The central measures of the study were the baseline and follow-up interviews, which had five components:

1. Demographics and background information on the relationship with the offender (at T2 – the current status of the relationship was asked).
2. History of abuse in the relationship and over the past six months (at T2 – abuse since the first interview).
3. Actions taken by the victim or the criminal justice or other systems that could mitigate her risk (repeated at T2).
4. Risk assessment (1 lethal risk and 1 recidivism risk instrument plus 19 "non-redundant questions" and self-perceived risk at T1 (at T2 - repeat of one of the two risk questionnaires administered at T1 plus four or five non-redundant items and self-perceived risk).
5. Impact of the questionnaire on perception of risk and self-protective actions.

The other major measure used was recidivism by the offender after the baseline interview, measured by arrest data provided by state criminal justice agencies. These measures are more fully described below.

Demographics and background on the relationship

In addition to the standard questions about age, race and ethnicity, and educational attainment of the offender and the victim, questions were added at T2 to learn more about the cultural background of the victim and offender. These questions were modeled on the US Census questions. At T1, we asked about country of birth of the victims. (We did not want to ask questions that could be interpreted by the victim as interest in her immigration status.) At T2, we asked country of origin of ancestors of US born participants, country of origin of the offender. and language spoken at home. In regard to the relationship, we ascertained legal status and level of current and past involvement.

History of abuse in the relationship

This information constituted our outcome variable at T2, but also provided a history of the severity and types of abuse. We used items from the Women's Experience of Battering (WEB; Hall-Smith, Smith & Earp, 1999) to assess the victim's psychological experience of abuse, items from the HARASS (Sheridan, 1998), to tap into stalking and harassment, and items from the Revised Conflict Tactics Scale (CTS2; Straus, Hamby et al., 1995) to assess physical and verbal abuse. Items from the CTS2 were used to construct a severity of injury index as an outcome measure, informed by the medical knowledge of the principal investigator.

The injury items used to construct the severity of injury index were drawn from the answers of the following question:

"*Please tell me whether you ever had any of the following injuries because of a fight with your partner." [For yes answers, follow up with: "Did it happen in the past six months."]*

Table C1. Injury Items

Physical pain that still hurt the next day
Sprain, bruise or small cut
Blacked out from being hit on the head
Broken bone
A permanent impairment or disability, such as loss of hearing, sight, or difficulty walking
Internal injuries to vital organs
Lost consciousness due to him choking you
Lost consciousness for more than an hour due to head injuries
Lost so much blood you had to receive a transfusion
Needed surgery to treat an injury from a fight
Been hospitalized or in rehab for more than four days because of injuries from your partner
Has he done anything that might have killed your or nearly killed you, whether or not he intended to?
Did you ever suffer a miscarriage because of a fight with your partner?
Did he try to kill you?

Severity of Abuse Outcome Variable

The injury items from the CTS2 and the stalking and threatening items from the HARASS instrument were used to construct a scale of severity of outcomes at T2 for analyzing the accuracy of predictions. This scale was not strictly ordinal, because, for example, stalking as an outcome could not be clearly assigned a level. Each item was assigned to one of the following outcome variable levels:

Table C2. Abuse Outcome Severity Categories (Primary Dependent Variable at follow-up)

None
Verbal abuse (e.g. name calling)
Psychological abuse and harassment (e.g. controlling behavior)
Stalking and threats
Physical abuse: Low (e.g., grabbing, slapping; causing bruises, small cuts)
Physical abuse: Medium (e.g., kick, slam up against a wall)
Physical abuse: High (e.g., strangling, sexual abuse with force, serious injury such as broken bone)
Physical abuse: Very High (e.g., use of a weapon, potentially life-threatening injury)

Each participant's outcome was assigned to the highest level of abuse she experienced during the period between baseline and follow-up interviews. Concurrent construct validity of the scale was supported by significant and moderately strong correlations with the CTS2 severe abuse with frequency weights variable ($r=.524$; $p<.000$), and with minor abuse (.602; $p < .000$), a summary score based on the WEB items ($r=.235$; $p<.000$), and a summary score based on the HARASS items ($r=.168$; $p<000$).

Protective actions

We were concerned about the possibility that a risk assessment instrument or method might predict a high level of risk but appear to be invalid when measured against outcomes because, although the risk was real, victims or the systems from which we recruited them took actions that mitigated that risk. To control for that possibility, we created questions to determine what protective actions had been taken by the victim or systems. This set of questions also gives us information about effective actions and interventions.

Questions about "self-protective actions" included securing weapons or items for self-defense, changing locks, and "going someplace where he couldn't find you." (See Table C3.) Participants were asked whether they had received counseling, and if so where, and whether they had adopted specific elements of safety planning usually recommended by victim assistance providers. They were asked about protection orders, reports to the police, and signed complaints. In addition to questions about whether the victim had gone somewhere inaccessible to her abuser, we also asked whether he had been someplace that made it impossible for him to get access to her (e.g., jail, out of the country, etc.).

Table C3. Self-Protective Actions and Interventions in Past 6 Months (Possible mitigation of risk)

PROTECTION ORDER/RESTRAINING ORDER – sought, received
ARREST for most recent incident
VICTIM SERVICES - safety planning, counseling, legal assistance, shelter, help with/for children; agency or other provider
NO CONTACT – LENGTH OF TIME: • Period when she was someplace where he couldn't find or see her (shelter, friend's/relative's/hotel, out of town, jail, hospital/other) • Period when he was someplace where he couldn't see her (jail, military, out of town, residential treatment program, hospital, other) • Period when they didn't see each other because one or both chose not to
MEDICAL TREATMENT for trauma or injuries
SAFETY PLANNING • Left documents, spare keys with friend or relative in case need to leave suddenly • Alerted family, friends, neighbors and/or children about violence • Cell phone, home alarm, lock change or other form of security
LAW ENFORCEMENT INVOLVEMENT - Called 911/reported to police/signed a complaint
SELF DEFENSE - Acquired mace, pepper spray or weapon
BATTERER PROGRAM - Abuser ordered to batterer program; if so, attended?

Finally, we asked if there had been a period when they hadn't seen each other because one or both of them chose not to have contact.

Analyses of risk assessment instruments, methods and questions were conducted with the outcome measures, and again controlling for the protective actions – particularly actions that should have resulted in the abuser being unable to contact the participant in person.

Risk Assessment

Participants were randomly assigned to receive either the Danger Assessment or a questionnaire we developed from the DV-MOSAIC areas of inquiry, and either the DVSI or the K-SID (random assignment is described later). In addition, all participants received another set of questions at T1 that do not constitute an instrument but that covered areas of possible risk prediction that were not included on any one of the four instruments (these came to be called the "non-redundant" items, meaning they were possible risk factors that did not appear on the instruments being tested). These questions were developed during the planning phase through a process of reviewing over 20 other risk assessment instruments or sets of questions currently in use and comparing them to the domains covered in the four sets of questions we were testing. From the pool of questions developed in this way, we selected those questions that had some basis in the research literature for having promise of being associated with dangerousness of a batterer.

Finally, we created questions on self-perceived risk, based in the literature that showed that victims are accurate in assessing lethal risk. We based these questions on Weisz, Tolman & Saunders (2000) in asking victims to rate the risk that their partner would abuse them in the next six months on a scale of one to ten. We first framed the question as the likelihood that the partner or ex-

partner would hurt them seriously. Many women answered in the affirmative, according to the interviewers, interpreting the question broadly, as meaning that he would do something hurtful, such as trying to take their children. We quickly added a second question on the likelihood that he would be physically abusive.

Impact of the questionnaire

As one of the purposes of risk assessment can be to raise the awareness of the victim about possible danger, we wanted to discover whether they felt that the interview process and answering the questions had changed their assessment of danger. To detect whether the questionnaire had an impact on their perception of risk, we asked the two questions on self-perceived risk at the beginning of the interview, immediately after the demographic questions, and again at the end of the interview. Also, at the end of the interview we asked in an open-ended question whether there was anything they thought they would do differently in the future as a result of answering the questions. In the follow-up interview, we asked the self-perceived risk questions at the end of the questionnaire, and asked if they had done anything differently since the previous interview as a result of participating in the study. Participants were given no information regarding their risk assessment scores.

Administration of Measures

Table C4 summarizes the major components of the baseline and follow-up interviews.

Table C4. Timing of Administration of Measures

Domain/Questions	Baseline	Follow-up
Demographics	√	Added questions on batterer race/ethnicity, victim language preference and origins
Relationship status	√	√
Self-perceived risk (1)	√	
Self-protective measures & interventions	√	√ Since last interview
History of abuse (WEB, Harassment & stalking, CTS, injuries)	√ ever & past 6 months	√ Since last interview
Lethal risk assessment	½ DA; ½ DV-MOSAIC*	¼ DA; ¼ DV-MOSAIC
Recidivism risk assessment	½ K-SID; ½ DVSI	¼ K-SID; ¼ DVSI
Non-redundant risk assessment items	1-19	¼ 4 or 5 questions
Self-perceived risk (repeated at end of interview)	√	√
Impact of interview on thinking/plans	√	Retrospectively – Did you do anything differently?

* Questions based on the DV-MOSAIC areas of inquiry

Research Sites

New York City Sites - Safe Horizon (Chris O'Sullivan, co-PI)

Safe Horizon is the nation's largest victim assistance agency. A private non-profit, it operates in all five boroughs of New York City and provides a range of services, from concrete assistance and advocacy through short- and long-term counseling. Among the services offered to victims of domestic violence are the New York City Domestic Violence Hotline; emergency shelter and transitional housing; a reception center and advocacy in the city's five family courts; advocacy and counseling in the city's five criminal courts; community offices in each of the five boroughs, which provide counseling and support groups; and legal representation in family courts. The research project capitalized on and benefited from the good will that these services and Safe Horizon field staff have created, among clients as well as court personnel and judges. Safe Horizon program directors opened doors for us, made introductions, found interview rooms for us in their offices and solved problems.

Interview and Recruitment Sites

New York City Family Courts. Judge Joseph Lauria, the Chief Administrative Judge for the five New York City family courts endorsed the study and directed the supervising judges of each county's family court to cooperate in the study. Ultimately, arrangements were made to recruit from and conduct interviews in the Family Courts in Queens, Kings (Brooklyn) and the Bronx counties. Different procedures were approved by the Supervising Judges in these three courts.

In the Brooklyn, Queens and Bronx Family Courts, three part-time bilingual interviewers approached women who had just been granted an Order of Protection (OP) and interviewed them while they were waiting for the court order to be processed in the Court Action Processing (CAP) Unit, where petitioners pick up their orders. The Bronx and Brooklyn courts permitted the interviewer to sit in the court and hear the cases. Then they could approach women with qualifying cases after they left the court. In Queens, the interviewer had to approach women people waiting outside the CAP Unit – who might be waiting for a variety of other orders on other matters, such as adoption and custody.

Potential participants were screened first to ascertain that they were receiving an order against an intimate partner (father of their children, husband or ex-husband), and that they had experienced violence or threats of violence within the preceding six months. If the petitioner qualified and was interested in the study, she was usually taken to a private office in the court allocated to Safe Horizon. The interviewer would let the Court Officer know whom she was interviewing so that the woman could be alerted when her order was ready. Women interviewed in the court were read a consent form that differed in one respect from other consent forms: they were told that they would receive one payment of $50 for the follow-up interview six months later, rather than two payments of $20 and $30. The Court did not permit us to pay participants at the court, and they did not want the payment to occur while the case was still before the court. A total of 630 women recruited from the family courts completed baseline interviews, including 252 interviewed in the Kings County (Brooklyn) Family court, 213 in the Bronx Family court, and 165 in the Queens Family Court, by six interviewers fluent in English and Spanish.

NYC Domestic Violence Shelters. The second major source of participants in New York City was domestic violence shelters. Shelters provided a safe way for victims to participate. Safe Horizon operates domestic violence shelters and "safe dwellings" throughout New York City. The latter are free-standing apartments where one or two families may be housed; residents receive services at a nearby Safe Horizon community office. To recruit residents of shelters and safe dwellings, the bilingual NYC project director, sometimes accompanied by a bilingual interviewer, would attend the shelter's weekly residents' meeting at the shelter or community office and briefly present the study. Women would sign up, and be interviewed throughout the day after the meeting. They were paid $20 at the completion of the interview. In addition to six shelters operated by Safe Horizon, an emergency domestic violence shelter operated by another private non-profit, Volunteers of America, allowed us to recruit and interview residents for the study.

Women are referred to shelters in a different borough from their home borough for security reasons. Women were interviewed in three shelters in Brooklyn (65), one in Staten Island (37), one in the Bronx (44), one in Queens (21) and a new shelter that Safe Horizon opened in Manhattan late in the collection of baseline data (10). Altogether, 177 women residing in New York City domestic violence shelters were interviewed at baseline. In theory, the shelters keep information about where the women relocate when they leave shelter. All of the participants interviewed in shelters at baseline agreed to sign a waiver allowing the shelter to give us their new phone number and address when we were ready to re-contact them for the follow up study.

Safe Horizon Community Offices. Safe Horizon Community Offices in the five boroughs provide counseling, advocacy and referrals for victims of domestic violence. Attempts were made to recruit participants from among the community office clients, but these efforts were not so successful and efforts were redirected to the more productive sites described above – the family courts and shelters. The procedures for recruiting from the community offices included meeting with the staff to present the project and ask for referrals to the study, posting flyers, and an interviewer visiting and trying to recruit from the waiting room. The most effective method turned out to be individual counselors (or the community office director) referring clients to the project. The counselor would call the NYC project director with an interested client, and she would make an appointment to interview the participant at the community office or at the research office. Altogether, 11 women from Safe Horizon's Brooklyn, Queens, Staten Island and Bronx Community Offices were interviewed.

Johns Hopkins University Sites – New York City Hospitals (J. Campbell, PI)

Dr. Campbell established relationships with three public hospitals in New York City (Harlem Hospital Center in Manhattan; Kings County Hospital in Brooklyn; and Lincoln Hospital in the Bronx); a 4th hospital (St. Luke's-Roosevelt Hospital) was brought in during the last two weeks of baseline recruitment. In each hospital setting, protocols were established with hospital personnel to identify IPV victims coming to the attention of the Emergency Department and introduce those women not too acutely ill to the study.

The planned recruitment protocol needed considerable adjustment at each hospital. Some of the difficulties were attributable to personnel and procedural changes that occurred between the time the original protocol was written and the eventual start of the project (delayed due to the shutdown of Johns Hopkins's IRB then to the 9/11/01 terrorist attacks – see implementation

challenges). Other changes were accommodations to difficulties inherent in working in a crisis setting (e.g., long emergency room waits, such that women were exhausted and needed to leave by the time they came to our attention) with a low-income and/or transient population (e.g., women would agree to return, but then would not; women had no phone numbers to give us, and no person they trusted to give us as an alternate contact).

By the end of our recruitment, procedures across hospitals were quite different. For example, at Kings County Hospital, a staff research nurse conducted the interview in person, often on the same day the woman was seen in the hospital. At Harlem Hospital, the Director of the Center for Victim Support interviewed women who came to her office from various sources, including a support group for battered women. At Lincoln Hospital in the Bronx, all women were recruited from a battered women's support group and interviewed by a Johns Hopkins post-doctoral fellow, who is also an RN. At St. Luke's, the interviews were done on site by the post-doctoral fellow, or over the phone by the project director at Johns Hopkins (except for one interview in Spanish conducted by the Safe Horizon project director). While our expectations were that we would recruit 100 women per hospital setting, only 30 women total were recruited across all four hospitals.

Los Angeles Sites – Justice Research Center (Jan Roehl, co-PI)

Los Angeles County Sheriff's Department. In 1998, the Los Angeles County Sheriff's Department (LASD) implemented a special domestic violence response program in three of the department's 18 stations. Called STOP, for Safety Through Our Perseverance, the program was staffed by full-time deputies specially trained in domestic violence situations, who were to visit IPV victims at the scene of the incident if the STOP deputy was on duty and available then or within 48 hours of the incident if the STOP deputy could not respond to the 911. When on duty, the STOP deputies interviewed victims after regular deputies responded to the 911 call and arrested the offender or took other immediate action. In one LASD station, a victim advocate regularly rode with the STOP deputy. The STOP deputies' purpose was to provide greater safety for the victim, provide immediate victim assistance, and gather important evidence in a timely manner to improve prosecution. In 1999, when our working relationship with LASD began in earnest, the STOP deputies were interviewing a good portion of the IPV victims using the MOSAIC-20.

Data collection began in early 2002, with two full-time interviewers collecting incident reports from four stations (located in East L.A., Industry, Norwalk, and Pico Rivera) and the Altadena Police Department every week or two. Altadena Police Department is a small department near Pasadena, and was included at the request of the STOP commander who knew the Altadena Police Chief well. Office space was rented for the interviewers within a domestic violence advocacy organization centrally located in Pasadena. This arrangement provided the interviewers with safety, access to needed office services, and support from the DV advocates.

LASD ended the STOP program in January 2003 due to budget constraints (and grant support ending). JRC worked with domestic violence unit supervisors in the four STOP stations that were participating in the study at that time, and all agreed to continue to give us access to the incident reports through regular deputies. We were able to work out a similar arrangement with the Lakewood station, which did not have a STOP program. A handful of cases also entered the study from the Palmdale, Lancaster, and Carson STOP stations, through the central STOP office. The

Palmdale and Lancaster stations receive a large number of domestic violence calls, but the STOP program there was reluctant to participate in the study because they believed it would interfere with the calls the local domestic violence agency made to the victims.

Participants were contacted by phone and given the option of completing the interview by phone or in person. Nearly all preferred to be interviewed by phone. The cost in interviewer travel time and mileage to do in-person interviews was substantial in L.A. At the end of 2002, we closed the Pasadena office and continued the interviews with one-L.A. based interviewer, who continued to pick up the incident reports from the LASD stations. Baseline and follow-up interviews were conducted by phone by two interviewers based in JRC's office. This approach was more productive and economical than maintaining office space in the LA area and supporting full-time interviewers At all times, at least one of our interviewers was Spanish-speaking. We conducted 399 baseline interviews with women who had a case with the LASD.

Table C5. Baseline Interviews Completed by Site and Modality

Site	# Baseline Interviews	*Primary Modality*
NYC Family Courts	630	In person
LASD	399	Phone
NYC Shelters	177	In person
LA Shelter	60	In person
NYC Hospitals	30	In person (and phone)
NYC Community Offices	11	In person
Total	*1307*	*68% in person*

Los Angeles shelters. The L.A.-based interviewers also interviewed women staying at a shelter in Downey, California. The interviewers visited the shelter every 45 to 60 days (45 days was the maximum stay) and interviewed women in person after obtaining informed consent. Typically four to six women were interviewed each visit by each interviewer. A total of 60 interviews were conducted at the shelter.

Procedures

Although specific procedures varied from site to site, there were commonalties across sites that will be described here.

Training and Staffing

JRC developed a comprehensive *Interviewer's Manual*, which contained background information on the study, data collection procedures for the field sites, answers to frequently asked questions, and specific instructions for obtaining informed consent, conducting the interviews, ensuring victim safety, handling difficult situations, record keeping, and reporting to the central project staff. In January, 2002, the three PI's conducted a two-day training session in Pasadena. Attending the training were the two LA interviewers for JRC and the NYC project director and interviewer/translator. STOP deputies and their sergeant also attended the first part of this training, and Bob Martin, Vice President of Gavin de Becker Associates, visited the training and spoke about the philosophy behind the MOSAIC threat assessment method. Additional interviewers were

trained at JRC and Safe Horizon later in the project.

There was an overall project manager at Johns Hopkins, responsible for staying in touch with the interviewers at the other sites, dealing with the IRB at Johns Hopkins and the NYC hospitals, providing materials to the sites, and receiving the questionnaires and entering the data. Three people filled this position over the course of the project.

At Safe Horizon in New York City, there was a bilingual project director (MA/MSW) and three bilingual (English-Spanish) field interviewers, one for each court. The Project Director also did interviews in the shelters, as well as managing the staff, data, and payments to participants. The project went through two teams of interviewers and two project directors. One of the first team of interviewers stayed on in the office to conduct quality control and ongoing training of interviewers. One of the second team of field interviewers stayed on after follow-up calls were complete to manage the data and organize the criminal justice data search.

The Justice Research Center, which is based 300 miles north of Los Angeles in Pacific Grove, initially hired two fulltime Los Angeles-based interviewers, one of whom was fluent in Spanish. A project director at JRC oversaw data collection procedures, supervised the interviewers, developed a computerized instrument (subsequently unused) and managed the data flow.

Recruitment Procedures

Recruitment procedures for the smaller sites (shelters and hospitals) were described under those sites. Recruitment procedures for the two primary sites, NYC Family Courts and LASD cases, will be described below.

NYC Family Courts. Safe Horizon interviewers recruited participants from family courts in Brooklyn, the Bronx and Queens. The recruiting approach varied depending on factors such as the Supervising Judge's preferences, the skills of the interviewer, and the overall environment of the court. Generally, petitioners wait many hours outside the court waiting for their case to be called. Under our agreement with the court, we could not interview them before their court appearance, however. In some cases, the interviewer might screen the women, and then wait until they exited the court before recruiting them for the study. In the Bronx and Brooklyn Family Courts, interviewers would observe courtroom proceedings. Once the interviewer learned that the woman was there for domestic violence, she would exit the courtroom and wait for the petitioner outside (hearings on individual petitions typically last ten minutes). After the hearing, there is another long wait before the orders are written up in the Court Action Processing (CAP) Unit. At that point, the interviewers in Brooklyn and the Bronx approached the potential participant identified in court and invited her to participate in the study.

In addition, interviewers would recruit from the waiting area outside the CAP Unit (and this was the only method used in the Queens Family Court). To minimize sample bias, all women in the waiting area were invited to participate in the study and were screened. In order to ensure the safety of women, interviewers did not approach women who were sitting next to a man. Interviewers also used the Safe Horizon Reception Center in the court to recruit women. As with the recruitment of women in the waiting rooms, interviewers would ask everyone in the Reception

Center to participate in the study. Finally, flyers with the 1-800 number established for the study were posted in Safe Horizon Community Offices and the Family Court Reception Centers (for domestic violence victims). Women could call in to learn more about the study. Voice mail on the line identified the office as "Women's Health Survey."

LASD. After interviewers received the incident reports from LASD, they entered basic information into a database, in part to prevent women from entering the study twice due to additional incidents, and onto a Call Log, used to track the status of calls to the women. The interviewer then attempted to reach the woman by phone. The interviewers were instructed to make at least five calls to each potential participant, varying the time of day and day of week. In reality, many more calls were typically made. A huge number of cases received from the LASD had wrong, disconnected, or false phone numbers (interviewers called the "Time" lady more than once). Safety measures were implemented as well.

Interview Process

Phone interview procedures. These procedures were used at baseline for the LASD sample, and with all participants for the follow-up interviews. When a potential participant was reached, the interviewer first made sure it was a safe time to talk and then introduced her to the study. If she agreed to participate, her informed consent was obtained and a Contact Sheet was completed. The interview was conducted then if possible or an appointment was scheduled to do so as soon as possible. Each participant was offered $20 for the baseline interview and $30 for the follow-up interview. Money orders were sent to the address of her choice following each interview.

In-person interview procedures. The interviewer began by reading or summarizing the consent form to the participant. (In the courts, because the judges were concerned about payment to the participant, this section of the consent form was not read.) The participant and interviewer each signed two copies of the consent form, one for the participant and one for the research records. Participants were told that if they felt it was unsafe to keep the consent form with them, we would keep their copy and provide it to them whenever they wanted it.

Comparison of phone to in-person interviews. Conducting the in-person interviews in Family Court, interviewers felt they were able to recruit women who might have been reluctant to participate if they had been contacted by phone in a cold call, partly because of the long wait at the courthouse and partly because of the personal contact. Contrasting the baseline in-person interviews with the follow-up interviews by phone, the Safe Horizon staff found that, in the follow-up interviews, the women were less apt to cry (perhaps a function of time as well as modality), and did not disclose as much. They tended to give monosyllabic answers and run through the interview in a more methodical way. There were fewer distractions on the telephone, from children and other activities in the court.

Offender and Contact Information. The second step of the interview was getting information on the offender and contact information for the victim. Information collected on the offender included name and aliases, address, date of birth, and, if she knew it, Social Security number. In Los Angeles, the LASD incident reports contained this offender information, plus a driver's license number and other demographic information (race, for example).

Contact information collected from the participants included a mailing address (to which reminder letters for the second interview and postal money orders would be sent), their own phone numbers, and the "number of someone who will always know where you are." Quite a few women said there was no one.. When initial interviews produced no contact numbers, interviewers were trained to prompt participants – "Is there a friend; relative - mother or sister or aunt; a neighbor?" The prompt was often effective in producing a number.

In New York, the women would sometimes name a case manager. To accommodate these responses, a confidentiality waiver form was produced that could be presented to the case manager giving permission to give contact information to the researchers. The women were encouraged to tell this "alternate contact" that they were participating in a study and that it would be okay to give us contact information when we called. It is our impression that alternate contacts usually were not informed about the study.

We offered the women a cover story that we would use when we called back – "women's health survey" – and asked if that would be okay or if they preferred a different cover story. Most women agreed, except for those who said that no cover story was necessary. (In the second year of data collection, Safe Horizon established an 800 toll free number. Women were encouraged to call in to the number if they moved or got a new number.)

Random Assignment. Each participant received the main interview plus one of the two lethal risk assessment instruments/methods and one of the two instruments designed to assess risk of re-assault. The order of the risk assessment measures was counterbalanced, resulting in eight T1 interview conditions, as listed in the table below.

Table C6 – Time 1 Assessment Pair

Condition	T1 Assessment Pair	
1	K-SID	DV-MOSAIC
2	DV-MOSAIC	K-SID
3	K-SID	DA
4	DA	K-SID
5	DVSI	DV-MOSAIC
6	DV-MOSAIC	DVSI
7	DVSI	DA
8	DA	DVSI

When each interviewer started, she was assigned a randomly generated number (sometimes using a computer program, sometimes pulling a slip of paper out of a hat) between one and eight. This number would determine the condition assigned to the first participant she interviewed. From that point on, she would simply move to the next highest interview condition with each subsequent interview. This procedure resulted in an evenly distributed number of interviews per condition in the T1 data set, as shown below.

Table C7. Condition (Assessment Pair) Frequency in Data Set

Condition	Frequency in data set	Percent of T1 Interviews
1	163	12.5%
2	161	12.3%
3	165	12.6%
4	165	12.6%
5	161	12.3%
6	159	12.2%
7	161	12.3%
8	170	13.0%

Next, the interview was conducted, in English or Spanish, beginning with demographic questions. Interviewers reported that some women cried during the interview. The interviewer always stopped and suggested a break. Most of the women wanted to continue even though the recollection was clearly upsetting. Many said they appreciated being able to tell someone the whole story.

Depending on the interviewer, the participant, and the condition, the baseline interview took from 35 minutes to an hour. Interviews that contained the questionnaire based on the DV-MOSAIC method took 15 to 25 minutes longer than the others; it was the longest set of questions by far, and the questions were more complex and required more thought on the part of the participant and the interviewer. Choosing the response option that best matched the initial answer could require further exploration or reading the answer back to the participant. Especially in that condition, but in others as well, some participants became impatient toward the end of the interview and did not want to complete the interview because of concerns about time. In such cases, the interviewer tried to set an appointment to complete the interview by phone at another time. In a number of cases, we were never able to reach the participant and we had not progressed sufficiently in the interview to collect usable data – often we had only the consent form and contact information. For this reason, twenty-eight cases from the court were dropped from the sample.

At the end of in-person interviews, the interviewer thanked the participant and gave her referral information to obtain services. Women interviewed in person outside of the court settings or by telephone were paid $20 for the baseline interview. Women interviewed in the family court were given a round-trip metrocard for subway or bus travel, with a value of $4.

Follow up interviews

Timeline. Follow-up interviews were started in December 2002, and ended in early January 2004. The length of follow-up periods varied (and were controlled for in the analysis) for several reasons. Intensive baseline data collection continued until May 2003, and the staff could not keep up with the T2 interviews as they came due. In many cases, interviewers began trying to contact a participant by phone after six months had passed but couldn't reach her, and they continued to try to reach her over the following several months. Retrying numbers that were out of service at the initial call was sometimes fruitful as well.

Conditions at T2 and maintenance of random assignment. The T2 interview consisted of only four conditions, as each participant received the main interview plus only one risk assessment instrument or method. In order to keep the second interview as short as possible, non-redundant items were also divided among the four conditions, as detailed in the table below.

Table C8. Follow-up Interview Conditions

Condition at T1	Risk instrument/method repeated at follow-up	Non-redundant items
1 or 6	DV-MOSAIC	L1-L5
2 or 3	K-SID	L6-L9
4 or 7	DA	L10-L14
5 or 8	DVSI	L15-L19

Participants remained evenly distributed across interview conditions at T2, as shown below.

Table C9. Maintenance of Random Assignment at T2

Condition at T1	Condition at T2	# T2 Interviews Completed	Percent of completed T2's
1 & 6	1	194	25.2%
2 & 3	2	198	26.7%
4 & 7	3	187	24.3%
5 & 8	4	190	24.7%

Protocol for T2 Interviews. Follow-up interviews were conducted over the phone, except on the rare occasions when a participant requested to do the interview in person. Before contacting a participant for a follow-up interview, the interviewer checked the contact sheet for any instructions about safety (cover story, good times to call, etc.). Interviewers followed a standard protocol depending on whether a man, woman, or child answered the phone. Unless an interviewer was speaking directly to a participant, she never revealed the name of the study or the nature of the research. When someone other than the participant answered the phone and asked why the interviewer was calling the participant, the interviewer simply said that she was conducting a survey on women's health and could speak to the participant only. Interviewers did not leave messages on answering machines unless they had previously been instructed to do so by the participant. Interviewers recorded the outcome of each attempt made and any relevant details of the attempt (e.g., best time to call participant, reluctance of the person who answered to provide any information, tenor of response) on the contact sheet for future reference.

Upon reaching a participant, interviewers gave a brief introduction explaining who they were and why they were calling and then proceeded to ask the following three questions: (1) "Is this a good time to talk for a few minutes?" (2) "Are you able to talk privately at this time?" and (3) "Is this a safe time for you to talk, a time when you will not be overheard or interrupted?" (These same questions were asked on all initial phone contacts with the LASD sample.) If the participant answered "No" to any of these questions, the interviewer told the participant that she would call her back another time and asked the participant what time would be best.

Before beginning the interview, interviewers informed participants that the interview would last approximately 20 to 30 minutes and that they would receive a $30 money order ($50 if recruited from the court) after the interview was complete. Interviewers also reminded participants that their participation was entirely confidential, that they did not have to answer any questions that they did not wish to, and that they were free to end the interview at any time.

Before beginning the interview, interviewers reviewed both the participant's and the offender's information on the contact sheet in order to confirm its accuracy and to record any new information that the participant could supply. Interviewers were asked to probe for the offender's date of birth and social security number if they were not originally provided, as well as for any change of address of the offender and the participant.

Difficult to reach participants, delays and attrition. The major reason for attrition at T2 was our inability to reach the victim. We are not aware of any situations in which participating in the study caused serious problems for participants or put them in danger.

If they determined that the phone contact information for a participant was no longer viable, interviewers attempted to reach the alternate contacts by phone. On reaching someone at the alternate contact number, the interviewer asked for the alternate contact by name. If the named contact answered, the interviewer explained that the participant had given us the contact's name and number so that we could find out how to reach her if her number changed. Some alternate contacts readily gave the information, others required more explanation, and some simply said they would pass along the message. In a few cases, the participant was there, or called back from the contact's number at a later time. There were a number of cases in which the alternate did not know where the participant was, and cases in which the alternate's phone number was also disconnected. (See Table C11 - Reasons for Declining.)

If neither a participant nor her alternate contacts could be reached by phone after several attempts, letters were sent to the participant's address asking her to call Safe Horizon or JRC to schedule a time to do the interview. Occasionally these letters were returned by the post office with no forwarding address. Letters were then sent to an alternate contact's address asking them to pass the information along to the participant. In NYC, 209 letters were sent, and 30 women (14% of the letter recipients) called into the 800-number in response. (This information was not tracked for the LA sample.)

As interviewers systematically moved through all possible avenues of communication, proceeding from the most direct and immediate form of contact (calls to participant) to the least direct and immediate (letters to alternates' addresses), often several months would pass until the participant was finally reached to complete the interview.

Women living at a shelter at the time of the baseline interview were the most difficult to reach for the follow-up interview. When we contacted the shelter, we often found that the participant had moved. In these situations, interviewers tried to determine who the caseworker had been and to speak to her to see if she had any forwarding information for the participant. We were often told that the participant occasionally speaks with the caseworker, and that the caseworker would give our phone number to the participant the next time she called. Weeks or months could

pass until we received a phone call from the participant. In other instances, participants' and/or alternates' phone numbers were temporarily disconnected (sometimes for two or three months or more), but we would continue to call to check on the status of the service. We were sometimes able to get through to a participant many months later, when her phone service was reinstated. Because of interviewers' persistence and reluctance to make the determination that the final status of the participant was "unreachable," we continued to contact participants through the final weeks of the study; hence, some T2 interviews were completed after a year or more had elapsed after a baseline interview, and one was completed almost two years after.

Retention Rates

There were 1307 eligible and usable T1 interviews completed, and of these cases, 782 completed a T2 interview, for a retention rate of 60%. Retention rates varied by site, as detailed in Table C10.

A very small number of women declined to participate at T2. As can be seen in the table below, only a total of 23 women were reached and subsequently declined. Most non-responses at T2 were due to invalid contact information, or the inability to get the participant on the phone. Our re-contact procedure precluded us from leaving messages on answering machines or with any person other than the named alternate. Thus, it was not uncommon for our interviewers to attempt to reach a participant 10 to 15 times, each time reaching someone in the home and being told to call back later, but never reaching the participant. There were also occasions where we did reach the participant, and she set up an interview time for us to call her back, but she was not available during the call back time, or any time thereafter. The table below details the major categories of attrition in the study, by study site.

Table C10. Retention Rate at T2 by Site

Recruitment Site	T2 Retention Rate
NYC Family Courts	61%*
LASD	69%
NYC Shelters	45%
LA Shelter	33%
NYC Hospitals	47%
NYC SH Community offices	64%
Total	*60%*

*Retention rates in Family Courts ranged from 70% in Brooklyn, which also had the highest recruitment rate, to 52% in Queens Family Court, which had the lowest recruitment rate of the three courts.

Table C11. Reasons for Attrition at T2: Frequency and Percent of Baseline Sample

Reason for attrition	Safe Horizon (NYC Courts Shelters) N (% of T1) T1 N=818	JRC (LASD, LA Shelters) N (% of T1) T1 N=459	JHU (NYC hospitals) N (% of T1) T1 N=30
Retention rate	58%	64%	47%
1. All contact info invalid	194 (24%)	126 (26%)	6 (20%)
2. Valid contact info for participant but couldn't reach directly/never started interview	135 (17%)	31 (7%)	10 (33%)
3. Participant reached but declined	17 (2%)	6 (1%)	0
Total with baseline but no T2 interview	*346 (42%)*	*163 (36%)*	*16 (53%)*

1. All contact information invalid includes: disconnected phones, wrong numbers, letters returned, participant left shelter and left no forwarding information.
2. Valid contact information but didn't reach/interview includes: only reached an answering machine; ring/no pickup; being told to call back; leaving call-back number with the alternate who says she will give it to participant; speaking with participant but never a good time or unavailable at scheduled time.
3. Participant declines. Interviewer spoke to participant, who decided she no longer wanted to participate.

For the small number who declined (23) at Time 2, we do not have records of the reason for declining. Anecdotally, the interviewers occasionally passed along the reason given by the participant. These included a lack of time, that the relationship was fine, and that the woman had moved on in her life and did not want to dwell on the past.

Criminal Justice Data

California criminal justice recidivism (arrest for DV) data were obtained from the California Office of Criminal Justice Planning, through a search of the CLETS database. The search was done using the offender's name, gender, date of birth, social security number, and driver's license number. We had the date of birth for nearly every offender, the license number for the majority of offenders for whom we had incident reports, and a few social security numbers. No criminal history record was found for 92 offenders, 20% of the L.A. cases.

New York criminal justice recidivism (arrest for DV) data were obtained under a contract with the New York State Division of Criminal Justice Services (DCJS). The basis of the search is the same as in California, except that driver's license numbers are not used. Offenders who have

ever been arrested are best identified through a "NYSID number" (New York State ID), but we did not have NYSID numbers because of our data sources. Out of 856 names submitted to DCJS, 579 (68%) had a possible match. This rate is surprisingly high, as we recruited most of the victims from a civil court, and none was recruited from a criminal justice site.

Implementation Challenges

Delays

The Spring of 2001 was spent lining up site agreements. The two largest recruitment sites, the New York Family Courts and the Los Angeles Sheriff's Department, required major changes, as detailed in those sections. (Recruitment at the hospitals was delayed even further, for other reasons, and therefore less affected by the initial delays.) Safe Horizon and JRC hired interviewers in New York and Los Angeles, and Johns Hopkins staff submitted the protocol to their IRB.

Several unforeseen tragedies occurred. In the summer of 2001, all research projects at Johns Hopkins University and approved by the IRB were halted because of the death of a volunteer in a medical experiment. Our application was not processed until the IRB was reconstituted with new regulations. The requirements were more stringent, but the main problem was that we were delayed into the fall.

As our application was being reviewed by the new Johns Hopkins IRB in September, 2001, the attacks on New York City and the Pentagon occurred. Safe Horizon's main office, sevenblocks from the World Trade Center was closed from September 11 until December 9. There was no telephone or e-mail service for three months, and the NYC PI was redeployed fulltime to 9/11 related duties until December. The Risk Assessment project began in earnest in early 2002.

Another delay was related to negotiations with Gavin de Becker & Associates and changes in the MOSAIC application. The MOSAIC-20, which we had obtained and been trained on in the course of the planning study, had been replaced by the DV-MOSAIC, a more complex system that was not so readily adaptable to the purpose and design of this study. The MOSAIC-20, as used by the LASD, was an interview designed to be administered to the victim as soon as possible after a violent incident. The new DV-MOSAIC was not designed as an "instrument" in the usual sense of the word. Rather, it is intended to be completed by an investigator after gathering information from the victim, offender, and existing records.

Difficulty implementing DV-MOSAIC in questionnaire format. In DV MOSIAC, there are areas of inquiry with graded responses that are guidelines designed to prompt more concrete questions and consolidation of information already collected by a criminal justice professional. Converting these complex domains of inquiry into direct questions that an interviewer could ask a victim, perhaps not a highly educated victim nor one who was familiar with criminal justice terminology, was a lengthy process. The investigators tried to remain true to the spirit of DV-MOSAIC while at the same time maintaining the independence of the evaluation. In addition, the response options did not have a linear relationship to the questions or areas of inquiry. They often introduced new dimensions, and the questions would have to be revised to capture those dimensions. Then the questionnaire had to be translated into Spanish and back translated into English to verify the translation. The first version we sent into the field with the interviewers in

March 2002 did not work; the interviewers reported that it was too cumbersome. We pulled them back and reworked it (and translated it into Spanish) and sent the interviewers out again. Much training was required on understanding the intent of the questions and how to capture responses.

Recruitment rate

In both New York and Los Angeles, the pace of recruitment of participants fell below expectations. In NYC, there was no one to interview on some court days and paperwork related to each interview consumed a large amount of time. Interviewers got both physically (losing their voice after four or five interviews) and emotionally exhausted. The time of a floating interviewer positioned to bring fresh paperwork to the other interviewers and collect their interviewers was taken up traveling hours roundtrip to each borough.

The NYC interviewers wrote the following report on the difficulties they faced in recruiting in the family courts.

> In the New York City Family Courts, many women were happy to participate in the study, but a large number declined to participate. Some had been in court so long that they just wanted to leave upon receiving their order of protection. While there was usually ample time to complete the interview while the protection order was being processed, many women did not feel up to participating because of headaches and fatigue. Some decided to leave the court before receiving their order. In addition, the family court atmosphere was not always conducive to conducting an interview. At times, the woman and the batterer and their children were all present. In this set of circumstances, a woman would often decline to participate. It may have been dangerous for her to discuss the history of the abuse with the abuser in such close proximity. Also, many women did not want the children to be privy to all of the details of the abuse. Several women, with children and without, were concerned about confidentiality. Even though the rules of confidentiality were clearly articulated, they thought it best not to participate. Many women that were eligible simply did not want to talk about it.

Calculations were similarly disparate from reality in Los Angeles, where there were two full time interviewers. With the number of domestic violence 911 calls that the LASD had reported, the assumption was that there would be ample potential participants. In reality, however, the participating stations did not produce the large volume of cases anticipated, and the contact information for victims was particularly bad. Various strategies were attempted to improve the pace of recruitment, including re-staffing, that had positive effects.

Both JRC and Safe Horizon found it easy to supplement the numbers with a shelter sample. On the one hand, this is a high-risk population and therefore appropriate for testing the validity of risk assessments, but on the other, retention rates are low because the women often relocate after leaving shelter. In one visit to a shelter, between four and seven interviews were guaranteed. Although the shelters assured us that they had contact information for women when they left shelter and that they stayed in touch with the former residents, in reality, they often did not have contact information (or accurate contact information) for former residents.

Johns Hopkins University staff experienced even greater challenges recruiting from the health care system. The NYC hospital emergency departments did not prove to be useful recruitment sites because of a combination of factors. Although routine screening for domestic violence was the policy in these emergency rooms and would have afforded sufficient numbers of abused women from among whom to recruit, these policies were not implemented uniformly across shifts and settings. As the literature has shown, unless there is constant attention to periodic training on domestic violence and appropriate screening and referrals for health care professionals, maintenance of system supports for screening and intervention for abused victims, and trust in the on site interventions for domestic violence, screening and identification of abuse victims decreases over time even when policies have been introduced. This decline seems to have occurred in these emergency departments. There were far fewer domestic violence victims being identified whom we could attempt to recruit than had been originally assured. In most of the hospitals, therefore, recruitment was done not through the emergency department but through the crime victims treatment program located in the social work department of the hospitals.

In summary, baseline interviews began in March 2002 and concluded in June 2003. Contacting the victims and recruiting them into the study proved more difficult and took much longer than we had anticipated, and the numbers of cases provided by the hospitals, LASD, and Family Courts were lower than anticipated. Follow-up interviews were started in December 2002 and continued through June 2003. The lag between baseline and follow-up, which averaged nine months instead of six months but took as long as a year to attempt, undoubtedly increased the attrition rate.

Staffing and impact on interviewers

Johns Hopkins had no personnel paid by the project on site at the hospitals, and had to depend primarily on hospital personnel who were not directly compensated by the research project. We tried to remedy this by having a post doctoral fellow spend at least one day a week in the hospitals recruiting and interviewing, but this step did not fully solve our problems. We therefore recruited few women from these settings and our attrition from those settings was even higher that that of our other settings.

A not unexpected problem experienced by both JRC and Safe Horizon was supervising field staff long distance. JRC is located 300 miles north of Los Angeles, and field staff were on their own most of the time. It was difficult to ascertain how time was being allocated and the exact reason for low productivity. When the operation was moved to Pacific Grove, with a single employee handling the logistics in Los Angeles, matters improved. Similarly, in New York City, it was not always possible to keep tabs on the interviewers. Most of the seven interviewers who worked in the field over the course of the project were highly self-motivated, hard working, and responsible, but there were a couple of short-lived hires who did not prove entirely reliable.

Another factor that we did not adequately plan for was the impact of doing the interviews on the interviewers, especially those working with victims in person. There was an emotional toll of hearing the stories and seeing the wounds, day after day, month after month. The second Safe Horizon project director was a social worker who was able to help the interviewers, and sending them to a workshop on vicarious trauma was also a help, they reported. The stress of trying to recruit exacerbated the strain. Breaks from interviewing had to be scheduled. There were

substantial individual differences both in how much exposure to victims' stories an interviewer could tolerate, and ability to recruit. Logically, these attributes co-varied. There was also a fine line that not all interviewers were able to navigate between effective recruiting and intrusion. Nonetheless, the most productive interviewer of the project also inspired attachment and loyalty from the participants, who seemed to see her as both a friend and an expert (although she was neither).

Another source of conflict for the interviewers in New York City was a need to feel that the information they produced would be used to help battered women and would be interpreted responsibly. As most of the NYC interviewers were Latina, they were especially concerned with how data on their community would be used. As one said, they used membership in that community to give the non-Latina PI's access and they wanted to be sure that the data would not be construed in a way that could be hurtful. In the LA cases, the Spanish-speaking interviewer noted that these interviews tended to be longer than those conducted in English, and she felt that often the interviews were the first time the women had spoken of the violence to others. This interpretation was corroborated by respondents' frequent inability or reluctance to provide alternative contact information, typically saying there was "no one" we could call who would always know where the participant was if we couldn't reach her for the follow-up interview.

Summary

As a consequence of these challenges, both recruitment and retention proved to be much more demanding, took more staff and supervisor time, and had lower success rates than we had anticipated. Generalizing from studies conducted in geographically smaller, less populous, and less complex municipalities may prove misleading when estimating recruitment and retention rates in major metropolitan areas. We were fortunate that NIJ granted us an extension and that we were finally able to recruit the 1307 participants needed for adequate statistical power.

D. RESULTS

Sample Retention and Characteristics

Of the 1307 study participants enrolled at baseline, we were able to obtain follow-up interviews with 782 (60%) at time 2. Dissimilarity between respondents and non-respondents to the follow-up survey heightens the probability of selection bias in findings. In this study, the participants lost to follow-up were demographically similar to those retained in the study with some exceptions (Table D1a). Unemployed women were most likely to be lost at follow-up. Hispanic women and homemakers were slightly but significantly more likely to be retained in the study.

Table D1a. Demographic characteristics at T1 (Baseline) of respondents and non-respondents to T2 interviews.

	All T1 Respondents (N=1307) n (%)	T1/T2 Respondents (N=782) n (%)	T2 Non-respondents (T1 only) (N=525) n (%)
Racial / Ethnic group **			
African descent /Black	363 (28%)	209 (27%)	154 (29%)
Latina /Hispanic	694 (53%)	444 (57%)	250 (48%)
European descent /White	130 (10%)	72 (9%)	58 (11%)
Other racial /ethnic groups	117 (9%)	55 (7%)	62 (12%)
Foreign born	497 (38%)	295 (38%)	202 (39%)
Education			
Less than high school	445 (34%)	262 (34%)	183 (35%)
High school diploma /GED	412 (32%)	260 (33%)	152 (29%)
Some college or vocational school	334 (26%)	186 (24%)	148 (28%)
BA/BS or college degree	115 (9%)	74 (10%)	41 (8%)
Employment status **			
Working full or part time	573 (44%)	367 (47%)	206 (39%)
Other	734 (56%)	415 (53%)	319 (61%)
Employment situation (check all that apply)			
Employed full time	394 (30%)	251 (32%)	143 (27%)
Employed part time	180 (14%)	117 (15%)	63 (12%)
Homemaker **	227 (17%)	155 (20%)	72 (14%)
Looking for work	165 (13%)	98 (13%)	67 (13%)
Unemployed ***	479 (37%)	252 (32%)	227 (43%)
Student	141 (11%)	83 (11%)	58 (11%)
Abuser's relationship to respondent			
Spouse /common law spouse	368 (28%)	230 (29%)	138 (27%)
Ex-husband /ex-common law husband	90 (7%)	47 (6%)	43 (8%)
Estranged husband	189 (15%)	117 (15%)	72 (14%)
Boyfriend	108 (8%)	77 (9%)	31 (6%)
Ex-boyfriend	541 (42%)	311 (40%)	230 (45%)
Involvement /Cohabitation at baseline			
Cohabitating	268 (20%)	180 (23%)	88 (17%)
Still involved but not cohabitating	49 (4%)	32 (4%)	17 (3%)
On again, off again	46 (4%)	29 (4%)	17 (3%)
Not involved or cohabitating	403 (72%)	541 (69%)	403 (77%)

** $p < .01$, *** $p < .001$

Note: Some categories add to more than 100% due to rounding. For each analysis the actual Ns may vary slightly from those listed at the top of the column, due to missing data.

As can be seen in Table D1b. below, the sample was severely abused at baseline, with 43% having experienced at least one CTS2 "severe" act three or more times in the past six months. Also, 45% had experienced a threat or assault with a weapon, choking or attempted murder within the previous 6 months.

Table D1b. Severity of Abuse Assessed at Baseline (T1)

Summary Severe Acts Reported	**All T1 respondents N (%)**
Experienced *all* of the following acts by this partner, at least once, ever: punched or hit with something that could hurt, choked, slammed up against wall, beat up, kicked	311 (23.7)
Experienced *all* of the following acts by this partner, at least once in the last 6 months: punched or hit with something that could hurt, choked, slammed up against wall, beat up, kicked	194 (14.8)
Experienced at least one CTS2 "severe" act 3 or more times in past 6 months (includes any of above acts or: using force to make have sex, using knife or gun, burned or scalded on purpose),	569 (43.4)
Partner used a knife or gun on her, choked her, or "tried to kill" her in past 6 months	589 (44.9)

In terms of severity of abuse, baseline participants lost to follow-up had similar mean scores on the risk assessment scales to those of participants providing interview data at time 2 (Table D2). However, those who could not be recontacted scored significantly higher at baseline on frequency of severe abuse on the CTS2, and were significantly more likely to rate themselves as at risk for reassault and serious harm during the next year.

Table D2. Mean T1 (baseline) risk assessment scores, victim's perceived risk, and severity of abuse (CTS2) for T2 respondents and non-respondents

T1 Risk Assessment Instrument/Method	**Possible Range (Actual range for all T1 Respondents)**	**T2 Respondents (Total N=782) (N per instrument/ method) Mean**	**T2 Non-Respondents (Total N=525) (N per instrument/ method) Mean**
DA point score	-3 to 37 (-1 to 36)	(N=400) 15.02	(N=266) 15.91
DV-MOSAIC rating	1 to 10 (3 to 9)	(N=382) 6.85	(N=259) 6.97
DVSI point score	0 to 40 (0 to 28)	(N=391) 8.60	(N=260) 8.65
KSID risk score	0 to 3 (0 to 3)	(N=391) 1.09	(N=267) 1.12
Victim rating of likelihood partner will physically abuse her in the next year	0 to 10 (0 to 10)	(N=775) 5.01	(N=517) 5.36*
Victim rating of likelihood partner will seriously hurt her in the next year	0 to 10 (0 to 10)	(N=771) 4.63	(N=510) 5.05*
CTS2 severe physical abuse, frequency weighted	0 to 42 (0 to 42)	(N=782) 6.82	(N=525) 8.41**

Reliability

Although each risk assessment instrument being evaluated did not necessarily seek to measure a unidimensional construct, and internal consistency is less relevant for risk assessment than for traditional instruments, we calculated Cronbach's alpha statistics for the DA, DVSI, and KSID in Table D3. The DA had the highest alpha (.66), followed by the DVSI (**"**=.61). The KSID displayed little internal consistency (**"**=.04). For the DA, internal consistency was relatively consistent across demographic subgroups. The DVSI had higher reliability among non-Hispanic Whites (**"**=.71) as compared with Hispanics (**"**=.57). The KSID demonstrated the lowest reliability among Black respondents (**"**= -.25) and among respondents for whom the interview was conducted in Spanish (**"**= -.12). It should be noted, however, that the instruments vary in length, and that number of items will affect reliabilities, in that there is likely to be more redundancy and therefore higher alphas for longer instruments. In the table below, the DA is the longest with 20 items; the DVSI has 12 and the K-SID has 10 items. Again, this measure of reliability is not particularly meaningful with risk assessment, which aims to capture multiple dimensions with the fewest questions possible.

Table D3. Reliability statistics for the DVSI, KSID, and Danger Assessment by selected respondent characteristics.

	DVSI	**KSID**	**Danger Assessment**
	alpha (N)	**Alpha (N)**	**alpha (N)**
All respondents	.61 (388)	.04 (381)	.66 (394)
African Amer.	.60 (102)	-.25 (104)	.61 (105)
Latina/Hispanic	.57 (221)	.11 (215)	.66 (221)
Non-Hispanic Whites	.71 (34)	.15 (38)	.65 (40)
Foreign born	.59 (148)	-.10 (143)	.61 (155)
Spanish interview	.54 (68)	-.12 (64)	.61 (76)

*Adjusted for distribution in sample

Internal consistency calculations are not presented for DV-MOSAIC, because, the items asked of women were derived from the "domains of inquiry" of that assessment method by the investigators. The term "reliability" is used in a different sense when applied to the DV-MOSAIC. The authors of DV-MOSAIC consider a DV-MOSAIC risk score (0-10) to be "unreliable" if the DV-MOSAIC IQ score (0-200) is less than 125. In other words, the score becomes less meaningful if there is insufficient information about the case. Within the study sample that received DV-MOSAIC (n=641), 2.7% (n=17) had IQ scores of less than 125 (Table D4). Low DV-MOSAIC IQ scores were more common among those who were administered the interview in Spanish than among those administered the interview in English.

Table D4. DV-MOSAIC IQ scores by selected respondent characteristics

Respondent Characteristic	Mean (SD)	% < 125 (n)
All respondents (n=641)	162.3 (16.6)	2.7 (17)
African American (n=174)	166.0* (13.9)	.6 (1)
Latina/Hispanic (n=350)	160.3 (17.6)	4.0 (14)
Non-Hispanic White (n=60)	161.3 (17.0)	1.7 (1)
Mixed/Other (n=54)	163.7 (15.6)	1.9 (1)
Foreign born (n=136)	161.2 (17.1)	2.9 (7)
U.S. born (n=240)	162.9 (16.3)	2.5 (10)
Spanish interview (n=56)	157.9 ** (17.5)	5.9 (6)
English interview (n=319)	163.1 (16.3)	2.0 (11)

* $p \leq .05$, ** $p \leq .01$, two-tailed test of independence

Concurrent Validity

Concurrent validity assesses the degree to which an instrument measures the construct or outcome of interest (e.g., IPV) based on correlation with instruments known to measure similar constructs, when data for both the instrument and the similar construct are gathered at the same point in time. The concurrent validity of the risk assessment instruments and DV-MOSAIC was evaluated based on the correlation between each instrument or method's risk score and baseline levels of (ex)partner abuse on the CTS2, and the correlation with victims' perceptions of the likelihood that their partner would harm them in the future. Each of the instruments/method was correlated with the frequency and severity of physical abuse, and all but K-SID were correlated with the frequency and severity of sexual abuse (Table D5). The extent of both physical and sexual abuse were most strongly correlated with DA scores (which contains two items similar to the CTS2, thereby inflating the correlation) and least strongly correlated with the K-SID. This same pattern held with the instruments'/method's correlations with victims' beliefs about the likelihood that their (ex)partner would seriously hurt them within the next year (Table D6). However, DV-MOSAIC had the strongest correlation between baseline risk assessment scores and respondents' perception of their (ex)partner's likelihood of being physically abusive (r=.450). K-SID scores were unrelated to victim's beliefs about their risk of future of assault by their (ex)partners.

Table D5. Correlations between T1 (baseline) risk assessment scores and measures of severe physical and sexual abuse at T1 as measured by the Revised Conflict Tactics Scale (CTS2) frequency weighted scores

	CTS2 Severe physical abuse, frequency weighted	**CTS2 Severe sexual abuse, frequency weighted [a]**
DA point score	.448 **	.295 **
DV-MOSAIC rating	.346 **	.251 **
DVSI point score	.293 **	.189 **
K-SID risk score[b]	.140**	.101**
Victim rating of likelihood partner will be physically abusive in the next year	.215**	.186**
Victim rating of likelihood partner will seriously hurt her in the next year	.269**	.215**

* $p \leq .05$, ** $p \leq .01$, two-tailed test of independence

[a] While the physical abuse item is a composite variable, the sexual abuse item is based on a single severe sexual abuse item in the CTS2 used in this study (sex without her consent using force). Frequency is factored into this variable.

[b] The risk score was used for both the continuous & categorical variable analyses for the KSID

Another way to estimate concurrent validity was the correlation of the instruments/method with the victims' perceptions of risk. The frequency distributions of the two items assessing victim perception, both asked on a scale of 1-10, are presented in Tables D6 and D7 below. As can be seen, approximately one quarter of the victims thought there was no chance that their partner would be physically abusive (23.7%) or seriously harm them (27.8%) in the next year. Close to the same percentage were certain that such would happen (23.2% and 22.3%) with the median at 5 for both and the mean 5.15 for physical abuse and 4.8 for serious harm.

Table D6. Responses to the item "How likely is it that your partner will be physically abusive with you in the next year?"

Likelihood it will happen	Rating	Frequency*	Percent
No chance	0	306	23.7
	1-4	211	16.3
	5	*177*	*13.7*
	6-9	297	23
I'm sure it will happen	10	301	23.3
Total		*1292*	*100.0*

* 15 women who responded "don't know" are not included in table

Table D7. Responses to the item "How likely is it that your partner will seriously hurt you in the next year?"

Likelihood it will happen	Rating	Frequency*	Percent
No chance	0	356	27.8
	1-4	225	17.6
	5	*160*	*12.5*
	6-9	254	19.8
I'm sure it will happen	10	286	22.3
Total		*1281*	*100.0*

* 26 women who responded "don't know" are not included in table

Table D8. Correlations between T1 (baseline) risk assessment scores and victims' perceived risk of abuse at beginning of interview

	How likely partner will be physically abusive in next year	**How likely partner will seriously hurt you in next year**
DA point score	.247 **	.344 **
DV-MOSAIC rating	.450 **	.465 **
DVSI point score	.248 **	.297 **
KSID risk score	.175**	.143**

- $p \leq .05$, ** $p \leq .01$, two-tailed test of independence

DV-MOSAIC had the strongest correlation with victims' perception of risk (.45 -.465) while the DA and the DVSI had the next highest correlations and the K-SID the lowest. The DA correlation was notably stronger (.344) for serious harm than for physical abuse (.247).

Abuse Occurring Between Baseline and Follow-up Interviews

Among the 782 study participants for whom we had both baseline and follow-up interview data, 243 (31.1%) reported experiencing some level of physical or sexual abuse by their (ex) partners during the follow-up period.

Table D9. Types and number of incidents of abuse experienced during the follow-up period (minor & severe abuse divided by bold line).

Type of abuse perpetrated by (ex-) partner	Any occurrence N (%)	1 time N (%)	2 times N (%)	3-5 times N (%)	6+ times N (%)
Threw something at her	95 (17.5)	24 (4.4)	20 (3.7)	31 (5.7)	20 (2.5)
Twisted her arm/hair	103 (19.0)	30 (5.5)	20 (3.7)	38 (7.0)	15 (2.8)
Slapped her	62 (11.6)	23 (4.3)	12 (2.3)	14 (2.6)	13 (2.4)
Grabbed her	134 (25.2)	43 (8.1)	22 (4.2)	37 (6.9)	32 (6.0)
Pushed/shoved her	152 (28.3)	44 (8.2)	27 (5.0)	44 (8.2)	37 (6.9)
Slammed her up against a wall	81 (15.2)	38 (7.1)	16 (3.0)	19 (3.6)	8 (1.5)
Insisted on sex (no force)	105 (19.7)	18 (3.4)	27 (5.1)	38 (7.1)	22 (4.1)
Punched her/hit her with something that could hurt	**73 (13.7)**	**28 (5.3)**	**18 (3.4)**	**14 (2.6)**	**13 (2.4)**
Kicked her	52 (9.8)	13 (2.5)	19 (3.6)	12 (2.3)	8 (1.5)
Burned or scalded her	6 (1.1)	3 (0.6)	1 (0.2)	2 (0.4)	0 (0.0)
Beat her up	69 (13.0)	30 (5.7)	15 (2.8)	14 (2.6)	10 (1.9)
"Choked" (strangled) her	71 (13.3)	44 (8.3)	10 (1.9)	10 (1.9)	7 (1.3)
Used gun or knife on her	25 (4.7)	11 (2.1)	7 (1.3)	6 (1.1)	1 (0.2)
Used force to make her have sex	43 (8.1)	14 (2.6)	10 (1.9)	12 (2.3)	7 (1.3)
Tried to kill her [1]	46 (8.8)	--	--	--	--

[1] Frequency not measured.

Although most women avoided physical or sexual abuse during the follow-up period, some experienced frequent and severe forms of abuse (Table D9). For example, 69 women (13.0%) reported being beaten up by their (ex)partner, 39 of whom reported multiple beatings. Forty-six women (8.8%) reported that their (ex)partners tried to kill them during the follow-up period; 71 (13.3%) reported being "choked" or strangled; and 25 (4.7%) had a gun or knife used against them. One woman from California died but we were not able to ascertain the reason for her death, and one perpetrator from NY was killed by police in the course of a violent incident not involving the victim.

In the original proposal, the research team collapsed the categories of abuse into 8 levels of ascending severity as shown in (Table D10) roughly paralleling the categories of severity suggested for use with the CTS2. The most common category of physical abuse was for acts that were severe enough to be potentially lethal, those in the Physical Abuse: Very High category (n=88, 11.3%) followed by minor forms of physical abuse (n=80, 10.2%). Harassment and psychological abuse without any physical abuse was as common as was physical abuse (n=240, 30.7%). Only 16.0% were free of any form of abuse including verbal abuse.

Table D10. Form of abuse experienced during follow-up period.

Form of abuse[1]	Items in Category	N (782)	%
None		125	16.0
Verbal	Call names	48	6.1
Psychological / harassment	Controlling behavior	240	30.7
Stalking / threats	Stalking/threats to harm	126	16.1
Physical Abuse: Low	Twist arm/hair, grab, push/shove, etc. Caused sprain, bruise, small cut.	80	10.2
Physical Abuse: Medium	Punch, kick, etc. Caused physical pain that still hurt the next day.	26	3.3
Physical Abuse: High	Choke, Burn, Beat up, etc. Serious Injury (Blacked out from being hit on the head, broken bone, etc.).	49	6.3
Physical Abuse: Very high	Use gun/ knife, try to kill, etc. Life threatening injury (Lost consciousness due to choking, etc.).	88	11.3

[1] Categories are mutually exclusive – only highest level of abuse experienced is reported

Bivariate Associations Between Baseline Risk Assessment Scores and Abuse During Follow-up Period

The following analyses examine how strongly baseline measures of risk are associated with subsequent abuse experienced during the follow-up period. We examined bivariate associations between each of the risk assessment scores and subsequent abuse in several ways. First, we calculated Pearson correlation coefficient for the association between the baseline risk assessment scores and our 8-level severity of abuse measure (Table D11). The DA had the highest correlation coefficient with subsequent abuse (r=.382), followed by DV-MOSAIC (r=.217),

DVSI (r=.199), and the K-SID Risk Score (r=.131). The correlations increased slightly for the DA, the DV-MOSAIC, and the DVSI when cases with no potential victim-partner contact during follow-up were excluded from the analysis. These were cases in which the victim was in shelter or had moved away and the partner did not know where she was, or the abuser was incarcerated or had moved far enough away that the victim felt there was no chance of contact.

Table D11. Correlations between T1 (baseline) risk assessment scores and severity of abuse during follow-up excluding cases with no potential exposure to abuser. [2]

	Pearson correlation coefficient
DA point score	.382**
DV-MOSAIC rating	.217**
DVSI point score	.199**
K-SID risk score	.131**

We also examined the correlation between the frequency and severity of physical and sexual abuse during the follow-up period. Only the DA (r=.246), and the DVSI (r=.174) were significantly associated with the frequency and severity of physical abuse during follow-up (Table D12). Only the DA (r=.141) and the DV-MOSAIC (r=.132) had statistically significant association between the frequency and severity of sexual abuse during the follow-up and baseline risk assessment (Table D12).

Table D12. Correlations between baseline risk assessment scores and CTS frequency weighted severe abuse scores.

	Severe physical abuse, frequency weighted	severe sexual abuse, frequency weighted
DA point score	.246**	.141**
DV-MOSAIC rating	.069	.132*
DVSI point score	.174**	.084
K-SID risk score	.020	.111

Each of the three risk assessment instruments being evaluated has specific cut-points for distinguishing victims at different levels of risk. The DV-MOSAIC, however, is meant to be used only as a 10 point rating scale. This continuous scale was used in all of the correlational analyses. Gavin de Becker argues for the increased precision of the ten point rating scale and points out compellingly, that cut points or ranges ignore the difference between a case at the lowest end of a range and the highest end of a range. However, for calculating the rest of the analyses (including sensitivity, specificity and positive predictive power), cut points are needed. We therefore based a cut point on de Becker's suggestions (1,2; 3,4; 5,6,7; 8,9,10) as well as our data. As de Becker predicted, we found that none of this previously abused sample scored in the 1-2 range of DV-MOSAIC. Using these cut-points (in the case of DV-MOSAIC artificially imposed), Tables D13a-13e display the categories or ratings of predicted risk in terms of the eight hierarchical categories of abuse in our abuse outcome measure (Table D10).

[2] 27 cases in which the victim indicated that either she was in shelter or some other location unknown to the abusive partner, or the partner was in jail or deported during the entire follow-up period.

Danger Assessment. The four categories of danger based on the DA are highly and significantly associated with the level or type of abuse during the follow-up (Table D13a). Among victims scoring in the "extreme danger" category at baseline, only 5.9% experienced either no abuse or only verbal abuse whereas about half of victims in the "variable danger" category experienced no abuse or only verbal abuse by their (ex)partner. One in five women in the extreme danger category experienced very high (and potentially lethal) abuse during the follow-up period, compared with one case (1.5%) in the "variable danger" category and 5.1% in the "increased danger" category. Among the 50 women who had been administered the DA and subsequently experienced potentially lethal abuse, 44 (88.0%) had scores in either the severe danger or extreme danger categories. The likelihood of being stalked or threatened yet avoiding physical assault increased from 5.9% in the "variable danger" category to "19.6% in the "extreme danger" category, but the increased (15.2%) and severe (12.5%) categories were not in order.

Table D13a. Level of abuse at T2 by risk category based on the Danger Assessment.

Abuse during follow-up	**Variable Danger (DA score 1-7) N=68 (17.0%)**		**Increased Danger (DA score 8-13) N=99 (24.7%)**		**Severe Danger (DA score14-17) N=80 (20.0%)**		**Extreme Danger (DA score 18-40) N=153 (38.3%)**	
	n	**(%)**	**N**	**(%)**	**N**	**(%)**	**n**	**(%)**
None	26	(38.2)	21	(21.2)	9	(11.3)	8	(5.2)
Verbal	12	(17.6)	3	(3.0)	7	(8.8)	1	(0.7)
Psychological / harassment	15	(22.1)	32	(32.3)	24	(30.0)	51	(33.3)
Stalking / threats	4	(5.9)	15	(15.2)	10	(12.5)	30	(19.6)
Physical Abuse: Low	8	(11.8)	10	(10.1)	8	(10.0)	13	(8.5)
Physical Abuse: Medium	0	(0)	6	(6.1)	0	(0.0)	6	(3.9)
Physical Abuse: High	2	(2.9)	7	(7.1)	10	(12.5)	12	(7.8)
Physical Abuse: Very high	1	(1.5)	5	(5.1)	12	(15.0)	32	(20.9)

χ^2 = 102.571, p<.001; Kendall's **J**-b = .289, p<.001

DV-MOSAIC. Scores on DV-MOSAIC were also significantly associated with level of abuse at follow-up. Subjects that scored in the highest rating (8 -10) category on DV-MOSAIC were twice as likely as those who had scores of below 8 to experience potentially lethal abuse during the follow-up period (14.9% vs. 7.3%) or stalking and threats (26.1% vs. 12.9%). Those with the highest ratings on DV-MOSAIC were only half as likely to experience no abuse or only verbal abuse (13.5% vs. 27.5%). Among the 38 women who were administered the DV-MOSAIC and subsequently experienced very high (and potentially lethal) violence, 20 (52.6%) had scores at the highest ratings.

Table D13b. Level of abuse at follow-up by baseline DV-MOSAIC ratings

Abuse during follow-up	MOSAIC Rating 3 - 4 N=23 (6.0%)		MOSAIC Rating 5 - 7 N=225 (58.9%)		MOSAIC Rating 8 - 10 N=134 (35.1%)	
	n	**(%)**	**n**	**(%)**	**n**	**(%)**
None	12	(52.2)	37	(16.4)	12	(9.0)
Verbal	1	(4.3)	18	(8.0)	6	(4.5)
Psychological / harassment	5	(21.7)	72	(32.0)	41	(30.6)
Stalking / threats	1	(4.3)	31	(13.8)	35	(26.1)
Physical Abuse: Low	3	(13.0)	29	(12.9)	9	(6.7)
Physical Abuse: Medium	0	(0.0)	10	(4.4)	4	(3.0)
Physical Abuse: High	1	(4.3)	10	(4.4)	7	(5.2)
Physical Abuse: Very high	0	(0.0)	18	(8.0)	20	(14.9)

$\chi^2 = 46.34$, $p<.000$

DVSI. The general pattern seen between DV-MOSAIC baseline ratings and subsequent severity of abuse was also observed for the DVSI (Table D13d).

Table D13c. Level of abuse at follow-up by baseline DVSI risk category.

Abuse during follow-up	Not High Risk (DVSI score 1-8) N=179 (45.8%)		High Risk (DVSI score=9-28) N=212 (54.2%)	
	n	(%)	n	(%)
None	37	(20.7)	18	(8.5)
Verbal	17	(9.5)	9	(4.2)
Psychological / harassment	54	(30.2)	68	(32.1)
Stalking / threats	12	(6.7)	50	(23.6)
Physical Abuse: Low	26	(14.5)	17	(8.0)
Physical Abuse: Medium	9	(5.0)	5	(2.4)
Physical Abuse: High	11	(6.1)	17	(8.0)
Physical Abuse: Very high	13	(7.3)	28	(13.2)

$\chi^2 = 41.23$, $df=7$, $p<.001$

A higher proportion of those in the high risk DVSI category were more likely than those not in the high-risk category to suffer potentially lethal forms of partner violence during the follow-up (13.2% vs. 7.3%), while more subjects in the not-high-risk category were more likely to escape harassment, psychological abuse, and physical abuse (30.2% vs. 12.7%). Among the 41 women who were administered the DVSI and subsequently experienced potentially lethal abuse 28 (68.3%) were in the DVSI high-risk category.

K-SID. There was no statistically significant association between the KSID risk levels measured at baseline and subsequent partner abuse (Table D13e).

Table D13d. Level of abuse at follow-up by baseline KSID risk category.

Abuse during follow-up	**Low Risk (KSID score 0) N=153 (39.1%)**		**Moderate Risk (KSID score 1) N=136 (34.8%)**		**High Risk (KSID score 2) N=12 (3.1%)**		**Very High Risk (KSID partner violated past TRO) N=90 (23.0%)**	
	n	**(%)**	**n**	**(%)**	**n**	**(%)**	**N**	**(%)**
None	36	(23.5)	26	(19.1)	2	(16.7)	6	(6.7)
Verbal	11	(7.2)	6	(4.4)	1	(8.3)	4	(4.4)
Psychological / harassment	45	(29.4)	44	(32.4)	2	(16.7)	27	(30.0)
Stalking / threats	21	(13.7)	20	(14.7)	2	(16.7)	21	(23.3)
Physical Abuse: Low	12	(7.8)	10	(7.4)	3	(25.0)	12	(13.3)
Physical Abuse: Medium	4	(2.6)	4	(2.9)	0	(0)	4	(4.4)
Physical Abuse: High	7	(4.6)	11	(8.1)	0	(0)	3	(3.3)
Physical Abuse: Very high	17	(11.1)	15	(11.0)	2	(16.7)	13	(14.4)

χ^2 = 25.94, df=21, p=.209; Kendall's **J**-b = .128, approximate p=.001

We also examined differences in the mean risk assessment scores (or risk category for the K-SID) for each of the risk assessment methods across the eight categories of abuse (Table D13e). Except for the KSID, there were statistically significant differences in the mean risk scores across the eight categories of abuse for each of the risk assessment methods. In general, women in the stalking/threats (but no physical or sexual abuse) and the very high abuse categories had the highest risk scores. The lowest risk scores were for women in the no abuse, verbal abuse only, and the minor physical or sexual abuse categories. The relative difference between the two categories with no physical abuse and the highest level of abuse severity was greatest for the Danger Assessment which produced a broader range of scores compared to the other instruments/method. In other words, the abuse categories varied most dramatically across the DA risk categories.

Table D13e. Mean risk assessment scores by Level of Abuse at Follow-up

Abuse during Follow-up	DA Mean (SD) *N*	DV-MOSAIC Mean (SD) *n*	DVSI Mean (SD) *n*	KSID Mean (SD) *N*
None	9.61 (6.66) *64*	6.07 (1.46) *61*	6.07 (4.36) *55*	.69 (.89) *70*
Verbal only	9.13 (6.59) *23*	6.48 (1.23) *25*	7.15 (4.17) *26*	.91 (1.15) *22*
Psychol/harassment	15.49 (6.99) *122*	6.98 (1.16) *118*	8.97 (4.81) *122*	1.09 (1.15) *118*
Threats/stalking	17.33 (6.62) *59*	7.42 (.84) *67*	11.05 (4.42) *62*	1.36 (1.25) *64*
Phys Abuse: low	14.33 (7.82) *39*	6.54 (1.23) *41*	7.12 (4.19) *43*	1.41 (1.26) *37*
Phys Abuse: medium	18.08 (7.68) *12*	6.64 (1.15) *14*	6.21 (4.84) *14*	1.33 (1.30) *12*
Phys Abuse: high	15.84 (5.39) *31*	6.94 (1.26) *18*	8.75 (3.84) *28*	.95 (.97) *21*
Phys Abuse: very high	20.08 (6.32) *50*	7.29 (.984) *38*	10.41 (5.49) *41*	1.23 (1.22) *47*
Overall mean *Total N*	15.02 (7.50) *400*	6.85 (1.239) *382*	8.60 (4.86) *391*	1.10 (1.16) *391*

Sensitivity, Specificity, and Positive Predictive Value

In order to assess the sensitivity, specificity and positive predictive power of the risk assessment instruments and method, we arrayed the data by instrument/method-specific risk categories and examined 1) whether the study participant experienced any physical or sexual abuse during the follow-up; and 2) whether the participant experienced a severe or potentially lethal assault (Table D14). Looking at the prediction of *any assault* during the follow-up, the only statistically significant associations with instrument/method-defined risk categories were for the DA risk categories. Victims' baseline-level rating of likelihood of reassault was also significantly associated with actual reassault experience. To make the DA categories more comparable to the high/low risk categorization based on victims' perceived risk, we compared the percentage reassaulted in the two lower risk groups (variable and increased danger, 23.6%) with the percentage reassaulted in the two high risk groups (high and extreme danger, 41.7%). The risk ratio of 1.8 that one obtains from this comparison of DA low and high-risk groups is similar to the risk ratio of 1.7 based on victim's high- vs. low rating of risk.

The DA and the DVSI risk categories were both significantly associated with *severe* or potentially lethal abuse during the follow-up period. The likelihood of experiencing a severe assault during the follow-up period increased from 3% (2 of 67) in the variable danger category of the DA to 30% in the extreme danger category. The most abrupt change, from 3% experiencing severe abuse to 16% experiencing severe abuse occurred between the variable (lowest category or score of –1 to +7) and increased danger (second lowest category, score of 8-13 out of possible 40) categories. Women in the high risk group on the DVSI were nearly twice as likely as those in the

low risk group to experience a severe assault by their (ex)partner during the follow-up period, 26% vs. 14%. The high-risk versus low-risk comparison based on victims' rating of the likelihood that their partner would seriously hurt them (25% vs. 13%) was similar to that observed for the DVSI. The association between the DV-MOSAIC high-risk categorization and the likelihood of severe assault during follow-up approached statistical significance (p=.06).

Table D14. Instrument/method-defined risk categories measured at baseline by physical or sexual assault victimization during follow up

Risk category determined at baseline by instrument/method	N	Any assault n (%)	Severe Assault n (%)
Danger Assessment			
Variable danger (-1 to 7)	67	11 (16.4)	2 (3.0)
Increased danger (8-13)	98	28 (28.6)	16 (16.3)
Severe danger (14-17)	79	30 (38.0)	19 (24.1)
Extreme danger (18-38)	144	63 (43.8)	47 (29.9)
Total	388	p < .001	p < .001
DV-MOSAIC			
Rating scores 3-4	23	4 (17.4)	1 (4.3)
Rating scores 5-7	219	67 (30.6)	31 (14.2)
Rating scores 8-10	126	40 (31.7)	27 (21.4)
Total	368	p=.317	p=.060
Domestic Violence Screening Instrument			
Not high risk 0-8	179	59 (30.5)	25 (13.8)
High risk 9-28	194	67 (33.5)	50 (25.3)
Total	373	p=.748	p=.004
K-SID			
Low risk 0	151	40 (26.5)	21 (13.9)
Moderate risk 1	131	40 (30.5)	25 (19.1)
High risk 2	11	5 (45.5)	2 (18.2)
Very high risk 3	90	32 (35.5)	16 (17.8)
Total	383	p=.336	p=.688
Victim's Perception of Risk	any sev.		
Low risk (1-4)	313 355	73 (23.3)	45 (12.7)
Medium risk (5)	93 87	34 (36.6)	18 (20.7)
High risk (6-10)	343 303	134 (39.1)	75 (24.8)
		p<.001	p<.001

The sensitivity, specificity, and positive predictive value based on the instrument/method-defined risk categories are presented in Table D15. The three categories of predicted increased risk for the DA provide a wide range of values for sensitivity and specificity. The sensitivity for the "increased danger" category is .917 for the DA. While the increased danger category for the DA provides the highest sensitivity of any of the instrument/method-defined risk groups, it also has the lowest specificity. The "severe danger" category of the DA provides the maximum value for combined sensitivity and specificity of any of the instruments' categories of increased risk [.704 (sensitivity) + .492 (specificity) = 1.196]. The "very high risk" category for the K-SID provided the highest specificity. Positive predictive value (PPV) is a function not only of the predictive validity of the instrument or method but also of the probability of the event being predicted. The PPV is the proportion of cases in a risk category that end up being reassaulted. The "extreme danger" category of the DA had the highest values of PPV of the instrument/method-defined risk categories (.438).

Table D15. Sensitivity, specificity, and positive predictive value of instrument/method-defined risk categorization at baseline and any physical or sexual assault during the follow-up.

	Sensitivity	Specificity	Positive Predictive Value
Danger Assessment			
Increased Danger	.917	.219	.377
Severe Danger	.704	.492	.417
Extreme Danger	.477	.684	.438
DV-MOSAIC			
ratings 5-7	.826	.074	.310
ratings 8-10	.360	.680	.317
Domestic Violence Screening Instrument			
High risk	.532	.486	.345
K-SID			
Moderate risk	.658	.417	.332
High risk	.316	.759	.366
Very high risk	.274	.782	.356
Victim's Perception of Risk			
Medium risk	.697	.472	.385
High risk	.556	.589	.391

The sensitivity of the instrument/method-defined risk categories for predicting severe and potentially lethal abuse was generally greater than was the case for predicting any physical or sexual abuse (Table D16). Again, the "increased danger" category of the DA had both the highest sensitivity and the lowest specificity. Comparing the sensitivity of victims' perceived risk of being seriously hurt by their (ex)partners against that of the risk assessment instrument or method categories, the DVSI is similar, the DV-MOSAIC is notably lower, all but the extreme category of the DA is higher, and all but the moderate risk category is lower for the K-SID. The specificity of

victims' perceived risk for serious assault was higher than that of the DVSI, higher than the elevated and high danger categories of the DA, and higher than the moderate risk category of the K-SID but lower than the very high risk category of the K-SID and the highest ratings of the DV-MOSAIC and extreme danger on the DA. The PPVs of the DA and DVSI were slightly greater than victim's perception of risk but none of the PPV's were particularly high, probably related to the infrequency of severe physical or sexual assault during follow-up.

Table D16. Sensitivity, specificity, and positive predictive value of instrument/method-defined risk categorization at baseline and severe physical or sexual assault during the follow-up.

	Sensitivity	Specificity	Positive Predictive Value
Danger Assessment			
Increased Danger	.975	.211	.282
Severe Danger	.775	.523	.278
Extreme Danger	.538	.672	.299
DV-MOSAIC			
ratings 5 - 7	.983	.071	.168
ratings 8 - 10	.458	.679	.214
Domestic Violence Screening Instrument			
High Risk	.667	.517	.258
K-SID			
Moderate risk	.672	.408	.185
High risk	.281	.740	.178
Very high risk	.250	.768	.178
Victim's Perception of Risk			
medium risk	.674	.511	.238
high risk	.543	.624	.248

We plotted receiver operator characteristic (ROC) curves to assess the ability of each instrument or method to distinguish between those who experienced abuse during the follow-up period throughout the continuum of baseline risk scores. Sensitivity and 1-minus-specificity are the coordinates of the ROC curves for each value of the risk assessment score. In contrast to the previous analyses, ROC curves do not depend on the cut-points selected for risk categories. If the risk scores do not enhance prediction of future abuse, the area under the ROC curve is not significantly greater than .500. We calculated these ROC curves excluding the 27 victims with no potential exposure to the abuser during the follow-up period. For the prediction of any physical or sexual abuse, the area under the ROC was not statistically different from .500 for any of the risk assessment instruments/methods except for the DA (.635) as well as for victims' perceived risk (.599) (Table D17).

The area under the ROC curve was generally larger for the outcome of severe abuse than for any physical/sexual abuse. The highest value for the area under the ROC was for the DA predicting the likelihood of severe abuse during the follow-up (.670, $p<.001$). The area under the ROC curve for predicting severe abuse was statistically significant for the DVSI (.597, $p<.05$) and DV-MOSAIC (.589, $p<.05$), but not for the K-SID. The DA was the only approach to risk

assessment that was higher than the victim's perception of risk for both any abuse and severe abuse at follow-up.

Table D17. Area under the ROC after excluding 27 victims with no potential exposure to abuser during follow-up.

	Any physical or sexual abuse	Severe physical or sexual abuse
Danger Assessment	.635 ***	.670 ***
DV-MOSAIC	.513	.589*
DVSI	.508	.597 **
K-SID	.551	.537
Victim's perception of risk	.599 ***	.610 ***

*** $p \leq .001$, * $p \leq .05$

Integration of Criminal Justice Data

The data described above are based on victim self-reports during follow-up interviews. We complemented our assessment of the association between risk assessment scores and categories with subsequent abuse by gathering arrest data on the index perpetrators/abusers who committed the abusive acts that led to victims' recruitment into the study. Only offenses committed subsequent to the baseline interview were considered. Relatively few (5%-7%) abusers were subsequently arrested for domestic violence during the follow-up period. With such a low prevalence of DV-related arrest, the power to detect statistically significant difference between the risk categories is diminished. As can be seen in Table D18, there was no significant association among the DA or DVSI risk categories or victims' perception of risk and subsequent DV arrests. There was a significant difference among DV-MOSAIC categories, however: DV arrests were more common in the lowest risk category, which had only 34 cases. Similarly, there was a statistically significant association between risk categories and DV arrests, driven by the high prevalence of arrest among 29 subjects in the high-risk category. The 147 subjects in the "very high risk" K-SID risk category actually had the lowest prevalence of DV arrest during follow-up. A similar pattern was observed between the instrument-method defined risk categories and abusers being arrested for any violent offense.

Table D18. Instrument-defined risk categories measured at baseline and arrest of abusers during follow-up excluding women with no potential exposure to their abusive partners.

Risk category determined at baseline by instrument	N	Domestic Assault n (row %)	Any Violent Offense n (row %)
Danger Assessment			
Variable danger	108	7 (6.5)	13 (12.0)
Increased danger	169	3 (1.8)	9 (5.3)
Severe danger	123	4 (3.3)	13 (10.6)
Extreme danger	266	17 (6.4)	31 (11.7)
Total	666	p=.097	p=.139
DV-MOSAIC			
Rating 3 – 4	34	6 (17.6)	6 (17.6)
Rating 5 - 7	370	24 (6.5)	46 (12.4)
Rating 8 -10	222	14 (6.3)	25 (11.3)
Total	626	p=.045	p=.569
Domestic Violence Screening Instrument			
Not high risk	300	13 (4.3)	25 (8.3)
High risk	351	25 (7.1)	45 (12.8)
Total	651	p=.130	p=.065
K-SID			
Low risk	252	12 (4.8)	19 (07.5)
Moderate risk	230	17 (7.4)	36 (15.7)
High risk	29	5 (17.2)	5 (17.2)
Very high risk	146	5 (3.4)	16 (11.0)
Total	658	p=.016	p=.037
Victim's Perceived Risk of Reassault			
Low risk	517	36 (7.0)	56 (10.8)
Medium risk	177	10 (5.6)	18 (10.2)
High risk	603	31 (5.1)	71 (11.8)
Total	1297	p=.431	p=.795

In addition to examining arrests as a separate outcome, we also combined the arrest data with the survey data on reassault to provide a more complete measure of abusers' reassault experience (Table D19). Situations in which subjects report no further abuse yet the abuser was arrested for domestic assault following the baseline interview could be due to under-reporting by subjects or abusers assaulting other intimate partners. The general pattern between instrument/method-defined risk categories and the likelihood of reassault during the follow-up

period observed with victim survey data was apparent when the survey data were combined with arrest data from criminal justice agencies. The prevalence of reassault increased with higher levels of risk categories for the DA, but was unrelated to categories derived from the other instruments/methods. Victims' perceptions of their own risk were also unrelated to this outcome measure. When the outcome was confined to severe assaults and/or DV arrest, the DVSI risk classifications also reached significance (28.9% in the high risk group vs. 18.4% in the low risk group).

Table D19. Instrument-defined risk categories at baseline and reassault experience during follow-up using both survey and arrest data excluding women with no possible contact with their abusers.

Risk category determined at baseline by instrument	N	Any Assault or DV arrest n (row %)	Severe Assault or DV arrest n (row %)
Danger Assessment			
Variable danger	67	15 (22.4)	7 (10.4)
Increased danger	98	29 (29.6)	17 (17.3)
Severe danger	79	31 (39.2)	21 (26.6)
Extreme danger	144	64 (44.4)	44 (30.6)
Total	388	p=.007	p=.005
DV-MOSAIC			
Rating 3 – 4	23	8 (34.8)	5 (6.6)
Rating 5 - 7	218	75 (34.4)	41 (18.8)
Rating 8 -10	126	41 (32.5)	30 (23.8)
Total	367	p=.935	p=.540
Domestic Violence Screening Instrument			
Not high risk	179	67 (37.4)	33 (18.4)
High risk	194	71 (36.6)	56 (28.9)
Total	373	p=.868	p=.018
K-SID			
Low risk	151	44 (29.1)	26 (17.2)
Moderate risk	131	44 (33.6)	30 (22.9)
High risk	11	5 (45.5)	2 (18.2)
Very high risk	89	32 (36.0)	18 (20.2)
Total	382	p=.540	p=.694
Victim's Perceived Risk of Reassault	any sev.		
Low risk	313 355	88 (28.1)	68 (19.2)
Medium risk	93 87	37 (39.8)	19 (21.8)
High risk	343 303	136 (39.7)	77 (25.4)
Total	749 735	p=.005	p=.155

Table D20 shows sensitivity, specificity and positive predictive values calculated for the combined outcome of victim reported reassault and arrest during the follow-up period. The data show slight

improvement in PPV for all approaches (including victims' predictions) over the values calculated with victim reported outcomes alone (shown in D15).

Table D20. Sensitivity, specificity, and positive predictive value of instrument-defined risk categories at baseline and victim reported physical/sexual assault or domestic violence arrest during the follow-up.

	Sensitivity		Specificity		Positive Predictive Power	
	Any	Severe	Any	Severe	Any	Severe
Danger Assessment						
Increased danger	.892	.921	.209	.201	.386	.255
Severe danger	.683	.730	.486	.472	.426	.291
Extreme danger	.460	.494	.679	.666	.444	.306
DV-MOSAIC						
Rating 5-7	.935	.934	.061	.062	.342	.187
Rating 8-10	.331	.395	.652	.672	.325	.238
DVSI						
High risk	.514	.629	.477	.514	.366	.289
K-SID						
Moderate risk	.648	.658	.415	.408	.349	.216
High risk	.296	.263	.752	.739	.366	.200
Very high risk	.256	.237	.775	.768	.356	.202
Victim's Perceived Risk						
Medium risk	.663	.585	.461	.494	.397	.246
High risk	.521	.470	.576	.611	.397	.254

Finally, we conducted an ROC curve analysis excluding the 27 victims with no potential exposure to the abuser during follow-up and using the Criminal Justice data as well as victim self report (Table D21). The DA was the only risk assessment approach that distinguished future abuse cases from non-abuse cases better than chance, however, the area under the ROC curve approached statistical significance for the DVSI prediction of severe reassaults (p=.055). Victim's expectations about their risk for reassault and serious injury were significantly better than chance, but the area under the ROC curves was less than that for the DA

Table D21. Area under the ROC curve after excluding 27 victims with no potential exposure to abuser during follow-up with Criminal Justice data.

	Any physical or sexual abuse during follow-up from victim self report &/or CJ data	Severe physical or sexual reassault from victim self report &/or CJ data
Danger Assessment	.613 ***	.628 ***
DV-MOSAIC	.474	.525
DVSI	.487	.567 (p=.055)
K-SID	.511	.523
Victim's perceived risk	.572 **	.551 *

*** $p \leq .001$, * $p \leq .05$

Association Between Risk Assessment Scores and Protective Measures Taken

The bivariate associations between baseline risk assessment scores and subsequent abuse described above may significantly understate the ability of the instruments/method to discern risk for future assault because the most dangerous abusers and circumstances are likely to prompt greater efforts to protect victims from future violence. For this reason, we examined associations between risk assessment scores and actions taken to protect victims from further abuse. Table D22a below shows that, except for K-SID, instrument/method-defined high-risk categories were associated with increased efforts of victims to escape their abusive ex-partners. For example, women who scored in the "extreme danger" category of the DA were over five times more likely than women in the "variable danger" category to go someplace where the abusers could not find them (40.5% in extreme danger vs. 7.4% in the variable danger category). Victims in the high-risk level based on their DV-MOSAIC score were twice as likely as women who scored at lower levels of risk to go someplace where their abusers could not find them (44.8% vs. 24.9%). Women in the high-risk categories for the DA, DV-MOSAIC, and the DVSI were more likely than other victims to go to a shelter for battered women. Nearly one of every five victims in the DV-MOSAIC highest level ratings went to a shelter in contrast to 1 in every 13 at the lower ratings.

While there was no association between risk category and the victim leaving town, the likelihood that the victim had no voluntary contact with her abusive ex-partner increased at higher levels of risk determined by the instruments/method at baseline (Table D22a). About half of the victims in the highest risk categories chose to have no contact with their abusive ex-partners during the follow-up period. Women in high-risk categories at baseline also tended to be more likely than other women during the follow-up period to receive counseling, do safety planning, change the locks on their doors, or obtain a weapon for protection, usually mace or pepper spray[3] (data not shown).

As predicted, the likelihood that the abuser was in jail during all or part of the follow-up period was, in general, higher in the instrument/method-derived high-risk categories (Table D22b). Almost one-third (28.8%) of abusers in the extreme danger category of the DA went to jail during the follow-up period compared with only 10.3% in the "increased danger" category. For both DV-MOSAIC and the DVSI, abusers in the high-risk category were more than twice as likely to go to jail during the follow-up period. There was no pattern in the relationship between abusers leaving town and risk category.

[3] Only 2 of 769 women reported that they obtained a firearm for protection.

Table D22a. Victims' efforts to limit contact with abuser by instrument/method-defined risk category at baseline.

	n	went where abuser couldn't find her n (%)	went to a shelter n (%)	left town n (%)	no (voluntary) contact n (%)
Danger Assessment					
Variable danger	68	5 (7.4)	0 (0)	2 (0)	9 (13.2)
Increased Danger	99	29 (29.3)	8 (8.1)	9 (9.1)	40 (40.4)
Severe Danger	80	22 (27.5)	9 (11.3)	4 (5.0)	35 (43.8)
Extreme Danger	153	62 (40.5)	23 (15.0)	14 (9.2)	80 (52.3)
	400	p < .001	p =.006	p = .284	p<.001
DV-MOSAIC					
3-4 rating	23	0 (0.0)	0 (0.0)	0 (0.0)	5 (21.7)
5-7 rating	225	56 (24.9)	18 (8.0)	13 (5.8)	94 (42.0)
8-10 rating	134	60 (44.8)	26 (19.4)	13 (9.7)	71 (53.0)
Total	382	p < .001	p=.001	p = .148	p=.010
DVSI					
Not high risk	176	45 25.6)	12 (6.8)	11 (6.3)	55 (31.3)
High risk	212	73 (34.4)	33 (15.6)	10 (4.7)	114 (53.8)
Total	388	p = .059	p=.007	p=.506	p<.001
K-SID					
Low risk	153	42 (27.5)	15 (9.8)	10 (6.5)	52 (34.2)
Moderate risk	136	39 (29.5)	11 (8.1)	10 (7.4)	59 (43.7)
High risk	12	2 (16.7)	1 (8.3)	1 (8.3)	6 (50.0)
Very high risk	90	34 (37.8)	12 (13.3)	8 (8.9)	48 (53.3)
	391	p = .241	p=.635	p=.925	p=.030

Table D22b. Abuser went someplace during follow-up where he could not see her by instrument/method-defined risk categories at baseline.

	n	abuser went someplace where he couldn't see her n (%)	abuser went to jail n (%)	abuser left town n (%)
Danger Assessment				
Variable danger	68	12 (17.6)	7 (10.3)	1 (1.5)
Increased Danger	99	29 (29.3)	15 (15.2)	7 (7.1)
Severe Danger	80	36 (45.0)	13 (16.2)	17 (21.3)
Extreme Danger	153	57 (37.3)	44 (28.8)	9 (5.9)
Total	400	p = .006	p=.002	p<.001
DV-MOSAIC				
3-4 rating	23	3 (13.0)	3 (13.0)	0 (0.0)
5-7 rating	225	66 (29.3)	28 (12.4)	25 (11.1)
8-10 rating	134	48 (35.8)	36 (26.9)	8 (6.0)
Total	382	p = .051	p=.001	p =.080
DVSI				
not high risk	176	44 (25.0)	18 (10.2)	14 (8.0)
high risk	212	80 (37.7)	54 (25.5)	18 (8.5)
Total	388	p = .007	p<.001	p=.848
K-SID				
low risk	153	32 (20.9)	14 (9.2)	11 (7.2)
moderate risk	136	50 (36.8)	27 (19.9)	19 (14.0)
high risk	11	8 (66.7)	7 (58.3)	0 (0)
very high risk	90	34 (37.8)	25 (27.8)	5 (5.6)
	391	p = .004	p < .001	p=.070

Regression Analyses

Given that baseline risk assessment scores and risk categories are associated with a wide range of factors intended to reduce victims' risk of reassault, we estimated the relationships between baseline risk assessment scores both before and after controlling for these factors. Specifically, the models included as covariates in the models the following conditions during the follow-up period: length of time victim was potentially at risk for reassault during follow-up; whether the victim had any voluntary contact with the abuser during the follow-up period; whether the victim went to a shelter; whether the victim received counseling; whether the victim changed the locks on her doors; whether the victim got an alarm; whether the victim got pepper spray or other weapons; and whether the abuser was incarcerated for some of the follow-up period. We estimated multinomial logistic regression models for an outcome variable that collapsed categories from our

8-level outcome variable into the following four categories: 1) no physical/sexual abuse, stalking or threatening; 2) stalking and/or threatening but no physical or sexual abuse; 3) minor or moderate physical/sexual abuse; and 4) severe abuse. The ability of each instrument or method to predict across the four categories of abuse after adjusting for potential confounders was determined by the difference in the –2 log likelihoods for the model that included each instrument/method risk score, and the reduced model that excludes the instrument/method scores. This difference follows a chi-square distribution – the higher the difference and the smaller the significance level, the greater the instrument or method's contribution to predicting victim's abuse experience during the follow-up. Because there is no standard scaling for the risk assessment instrument/method scores that would make regression coefficients comparable, we also present Wald statistics and significance levels that assess the instrument or method's ability to predict each specific category of abuse.

The findings relevant to the risk assessment instruments and method are summarized in Table D23. The DA enhanced the predictive power of the models more than the other two risk assessment instruments and DV-MOSAIC. The Wald statistics for the baseline DA's associations with subsequent severe or potentially lethal abuse as well as with minor or moderate abuse during the follow-up is several times greater than those for the other two instruments and DV-MOSAIC. The DA and the K-SID risk score had the only statistically significant Wald statistics for predicting minor to moderate levels of abuse. However, DV-MOSAIC had the highest Wald statistic for predicting subsequent stalking and threats. DV-MOSAIC and DVSI as well as the DA had statistically significant predictions for severe or potentially lethal abuse.

We added covariates measuring victims' baseline expectations about the likelihood that their abusive (ex)partners would seriously hurt them during the next 12 months to determine whether the instrument/method scores added predictive power over and above victims' expectations. Each of the formal risk assessment instruments/methods provided statistically significant improvement in the model beyond the victims' own assessment of the likelihood that their abusers would seriously hurt them.

Table D23. Summary of multinomial logistic regression models for abuse type during follow-up.

	-2 log likelihood for reduced model [a] (df)	signif.	Severe Abuse [b] Wald (signif.)	Minor/ Moderate Abuse Wald (signif.)	Stalking and Threatening Wald (signif.)
Danger Assessment	46.71 (3)	<.001	34.27 (<.001)	9.79 (.002)	16.78 (<.001)
DV-MOSAIC	26.08 (3)	<.001	5.05 (.025)	0.89 (.346)	19.41 (<.001)
Domestic Violence Screening Instrument	22.40 (3)	<.001	10.80 (.001)	0.14 (.907)	15.68 (<.001)
K-SID risk score	11.85	.008	0.59 (.442)	8.98 (.003)	5.55 (.018)
Victim perceived risk of severe assault	27.63 (3)	<.001	21.08 (<.001)	0.18 (.674)	5.60 (.018)

[a] The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model estimated by omitting an effect from the final model. For each chi-square test, there are 3 degrees of freedom.

[b] Forms of abuse including use of or threats with a knife or gun, beating up, "choking" (strangulation) with loss of consciousness, burning, forced sex, and injuries including broken bones, internal injuries to vital organs, loss of consciousness, severe blood loss.

Using methodology similar to that used by Heckert and Gondolf (2004), we enteredthe predicted probabilities generated by the binary logistic regression models with each of the risk assessment scores as predictors along with the other covariates described above to develop ROC curves and area under the curve statistics. In contrast to the area under the ROC curve statistics generated by the raw risk assessment scores in which only the DA had predictions greater than chance, all of the under the ROC curve statistics with the predicted probabilities were statistically significant at the .001 level (Table D24). However, only the DA and victims' perceived risk at baseline were significantly associated with the odds of experiencing any physical or sexual abuse in the regression models that produced the predicted probabilities. The situation of significant ROC curve statistics while risk assessment instrument/method coefficients were not statistically significant in the regression models also occurred for DV-MOSAIC and K-SID predicting severe abuse. The area under the ROC curve statistics for both outcomes (any abuse, severe abuse) were highest for the models using the DA.

Table D24. Area under the receiver operator characteristic (ROC) curve using predicted probabilities from logistic regression models with the risk assessment scores and potential confounders as predictors.

Baseline risk assessment predictor	Any physical or sexual abuse during follow-up	Severe physical or sexual abuse during follow-up
Danger Assessment	.674 ***	.687 ***
DV-MOSAIC	.583 *	.647 ***
DVSI	.595 *	.616 **
KSID risk	.606 ***	.622 **
Victim perceived risk	.619 ***	.619 ***

*** p ≤ .001, ** p≤.01, * p≤.05

Table 25 summarizes the ROC curve analyses with and without controlling for protective actions and with and without criminal justice data (combining data from D17, D21 and D24). As with the previous analysis, the DA has the highest area under the curve including (columns 1 and 2) or excluding criminal justice data (columns 3-6) and controlling for (columns 5-6) or not controlling for (columns 1-4) protective actions. The other methods perform differently under the different conditions, with DV-MOSAIC second highest with severe reassault, without criminal justice records and controlling for protective actions; the DVSI second for any re-assault, without criminal justice records and controlling for protective actions; and Victim Assessment of Risk second for conditions without controlling for protective actions.

Table D25. Summary Table - Comparative areas under the ROC curve with and without CJ data, and with and without controlling for protective actions taken

Method	Any re-assault (w/CJ data) (n=1307)	Severe re-assault (w/CJ data) (n=1307)	Any re-assault w/o CJ data (n=782)	Severe re-assault w/o CJ data (n=782)	Any re- assault w/o CJ data controlling for protective actions (n=782)	Severe re-assault w/o CJ data controlling for protective actions (n=782)
Danger Assessment	.613 ***	.628 ***	.635 ***	.670 ***	.674 ***	.687 ***
DV-MOSAIC	.474	.525	.513	.589*	.583 *	.647 ***
DVSI	.487	.567	.508	.597 **	.595 *	.616 **
K-SID	.511	.523	.516	.514	.606 ***	.622 **
Victim assessment of risk	.572 **	.551 *	.599 ***	.610 ***	.619 ***	.619 ***

* p ≤ .05, **p<.01, ***p<.001

The findings from the regression analyses described thus far indicate whether the risk assessment scores significantly improve the prediction of subsequent abuse by a violent partner better than by chance. While the likelihood ratio chi-square statistics and significance levels allow for comparisons across instruments/method, it is nevertheless difficult to interpret with these statistics alone how well each risk assessment score enhances prediction. To enhance

interpretability of the regression findings and to assess the utility of the cut-points derived by the instrument authors, we re-estimated the models using dummy variables to represent differences in the probability of reassault after baseline assessment across different categories of risk for each instrument or method after controlling for the effects of protective actions. The adjusted odds ratios derived from those analyses for the risk categories are presented in Table D25. These adjusted odds ratios can be interpreted as the odds of the specific category of abuse (e.g., severe abuse) occurring among women in the category relative to women in the reference category (e.g., odds in high risk group vs. odds in low risk group). The reference group in each analysis is the lowest risk category ("Variable Danger" for the DA and "Low Risk" for the K-SID).

As was the case when the risk assessment scores were analyzed as continuous variables, the DA provides the strongest predictive validity when categories of risk are assessed in the regressions (Table D26). The most dramatic effects for the DA are observed for predicting severe and potentially lethal forms of abuse: the odds were 17 times higher in the Extreme Danger category and 11 times higher in the Severe Danger category compared with the odds in the Variable Danger category. The ability of the DA categories to predict subsequent stalking and threatening behaviors was also impressive. For example, the odds that a victim would be stalked or threatened during the follow-up period were 6 times higher in the Extreme Danger category and nearly 3 times higher in the Severe Danger category than in the Variable Danger category. The data in Table D25 also reveal that the adjusted odds for experiencing severe or potentially lethal abuse were 6.5 times higher in the highest ratings groups for DV-MOSAIC than in the lowest. However, this difference was not statistically significant due to the very small number of cases (n=23) in the lowest risk category. The largest difference between DV-MOSAIC risk groups was observed for subsequent threats and stalking (with no physical or sexual abuse) where the odds were 11 times higher in the group with the highest rankings than in the group with the lowest rankings. High-risk status on the DVSI was associated with a 2.5-fold increased odds of severe abuse and a 4.7 increased odds of being stalked or threatened with violence. After controlling for the effects of protective strategies, K-SID risk categorization did not significantly enhance prediction of subsequent as measured by the 4-level abuse outcome variable. Nevertheless, the odds of experiencing minor-to-moderate level physical abuse and threats and stalking were significantly higher in the "very high risk" K-SID group compared to the "low risk" group. Based on the –2 log likelihoods for reduced models, categorization based upon victims' baseline self assessment of their risk of being seriously hurt was more accurate in predicting victims' status on the 4-level abuse outcome variable than was categorization based on the K-SID or DV-MOSAIC.

Table D26. Adjusted odds ratios and overall test of risk assessment categories on effects on abuse during the follow-up period.

	-2 log likelihood for reduced model [a] (df, significance)	Severe Abuse [b] aOR (signif.)	Minor/ Moderate Abuse aOR (signif.)	Stalking and Threatening aOR (signif.)
Danger Assessment	49.99			
Increased Danger	(9, p<.001)	3.7 (.056)	2.7 (.045)	2.7 (.103)
Severe Danger		11.0 (<.001)	2.1 (.200)	2.7 (.123)
Extreme Danger		17.2 (<.001)	12.7 (.004)	6.5 (.002)
DV-MOSAIC	21.35			
5-7 ratings	(6, p=.002)	3.4 (.247)	2.5 (.259)	4.2 (.175)
8-10 ratings		6.5 (.079)	2.2 (.157)	11.4 (.022)
Domestic Violence Screening Instrument	25.77 (3, p<.001)	2.5	1.1	4.7
High risk		(.003)	(.739)	(<.001)
K-SID	13.57			
Medium risk	(9, p=.138)	1.2 (.682)	1.2 (.655)	1.0 (.992)
High risk		1.6 (.603)	4.9 (.060)	2.0 (.454)
Very high risk		1.2 (.617)	3.2 (.007)	2.2 (.037)
Victim Perceived Risk	30.16			
Medium risk	(6, < .001)	2.1 (.035)	1.8 (.084)	1.9 (.059)
High risk		2.7 (<.001)	0.8 (.468)	1.8 (.013)

[a] The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model estimated by omitting an effect from the final model. For each chi-square test, there are 3 degrees of freedom.

[b] Forms of abuse included use of or threats with a knife or gun, beating up, "choking" (strangulation) with loss of consciousness, burning, forced sex, and injuries including broken bones, internal injuries to vital organs, loss of consciousness, and severe blood loss.

E. DISCUSSION

This study is a landmark prospective field trial of three instruments and one threat assessment system. It was complicated to implement and the results are complicated to interpret. We have considerable new information and exciting results from the trial, but one of the realities of the

science of risk assessments is that there are many ways to evaluate the validity of an instrument and the individual risk factors or domains of inquiry of a threat assessment system. Our results help in the evaluation of these strategies for assessing risk or threat in the area of domestic violence but do not give us a definitive answer as to what instrument or method a particular agency should be using, or the absolute utility of these methods.

Before we turn to a discussion of the results, it is important to understand both the strengths and limitations of the study. Among its strengths are the experience and expertise of the research team, the successful random assignment of the instruments and method, the large sample size, the multiple sources of participants in the study (courts, law enforcement, shelters, health care and victim services), the multiple sites in our two most populous states, the collaboration of the team with the criminal justice system and domestic violence service agencies in the locales studied, the instrumentation of the study, the linkage of victim and re-arrest data, and the in-depth, multiple approach analytic strategies employed including substantial controls for baseline danger and other self-protective steps taken by victims. Although the strengths are many, the limitations are also important to take into account. The attrition rate (40%) was disappointing in spite of careful attention to strategies suggested in the literature such as obtaining alternate contacts and obtaining the participants' permission to have these alternate contacts give the investigator new contact information. Our compensation for interviews was not particularly large, and perhaps if larger could have been more of an incentive. We were able to obtain additional participation by sending letters out to the safe address indicated by participants reminding them of the $30 compensation for the second interview ($50 for most of the NYC sample, those interviewed in the courts). We were able to raise the NYC participation rate from 54 to 57% with that strategy.. However, further pursuit of this strategy was hampered by the fact that we did not receive IRB approval to send the lettersuntil we were close to the time that the follow-up interviews had to end. We also found original recruitment to be far slower than anticipated due to a whole host of implementation issues and we were still enrolling participants for Time 1 interviews when we were initiating Time 2 follow-ups. This stretched already hard-pressed staff and resources further than was optimal for full attention to retention.

In spite of the relatively low retention rate, there were few significant differences between those who returned and those who did not: those who returned were less likely to be unemployed, more likely to be Hispanic and somewhat less severely abused than those that did not. Importantly, there were no significant differences in mean baseline risk assessment scores between those retained and those lost at follow-up. This finding strengthens the case that there was not systematic bias introduced by the attrition in the study, although the fact that those retained scored significantly lower on frequency of severe abuse raises some question. Furthermore, although our sample was large enough to carry out our major analyses with sufficient power, the attrition resulted in some subgroup analyses having empty cell sizes and lower than optimal power.

Another significant limitation was our inability to implement the DV-MOSAIC threat assessment system in the way it is intended to be used. Rather than a trained assessor using the DV-MOSAIC system as a guide for assessing risk after collecting information from the victim and the criminal justice records, the only way to use DV-MOSAIC in the study design was to translate the domains of inquiry and the multiple possible responses contained with the DV-MOSAIC format (with its embedded protective as well as risk factors) to a questionnaire format. This made for a long questionnaire. The DV-MOSAIC developers also caution that the 1-10 scoring is not meant to be a

ratio level scale (with statistically supported uniform intervals between points) but rather an easily understood indicator of escalating threat. However, for the statistical procedures we employed, we had to treat the scale as if it were ratio level. Our division of the scale into 4 levels was also our decision to enable the use of certain statistical procedures and enhance comparability with the divisions suggested for use by the other instruments, not what was intended by the DV-MOSAIC developers. The ROC curve analysis is probably the most accurate assessment of the threat assessment characteristics of the 10-point scale of DV-MOSAIC.

Readers of this report should note that the principal investigator on this study is the author of the DA. To guard against any possible bias in interpreting the findings, the two co-PI's, who have no professional stake in the DA and are appropriately questioning about DV risk assessment in general, were primarily responsible for the implementation of the study and carefully interrogated all results and their interpretation. The input of the Technical Work Group also has assisted in maintaining the objectivity of the final report.

The other issue with the DA is that we had proposed to evaluate the new version of the DA with its weighted scoring and levels of risk that has been developed as a result of the findings of the 11-city study (Campbell, Webster, Koziol-McLain et al., 2003). Unfortunately, when the baseline questionnaire was designed, there was one item about arrest about which a final decision had not yet been made as to whether to include it on the revised DA and how it would be worded. The arrest variable had turned out to be protective in the 11-city study which is the opposite of some of the work on reassault (e.g. Jackson et al., 2003) although consistent with reports of arrest as a deterrent at least for those with a stake in conformity (Maxwell, Garner & Fagan, 2001). We therefore were re-examining all the data related to this item and making sure it was a valid finding plus testing various wordings of the item in practice settings before making this decision. When we decided how to include the item in the revised DA, we had already collected the majority of Time 1 data and therefore only had prior DV arrest information on half the perpetrators, from the participants who were administered the K-SID. In most cases, including the arrest item as a protective item improves the validity of the DA but this information is not presented in this report in order to preserve the general comparability of the findings in terms of sample size.

Finally, although we have completed extensive analyses, analysis is not yet complete. Additional analyses will be completed on this data set for peer reviewed publications over the next few years. We particularly still have work to do in examining the predictive value of the other potential risk factors we included and analysis of individual items. Modification of current risk assessment approaches and development of others can be informed by our findings in that realm.

Discussion of Findings

Reliability. Although internal consistency is arguably not an appropriate standard for risk assessment instruments since a) they are intended to combine independent risk factors rather than measure a one-dimensional construct, b) they aim at brevity, and c) they typically use dichotomous responses. These characteristics all tend to lower alpha coefficients. Nonetheless, the DA and the DVSI both achieved respectable internal consistency overall (Standardized Alpha = .76 and .63 respectively) within most of the ethnic groups. The DVSI was less internally consistent among Spanish speaking and/or foreign born participants and is also shorter (12 items) than the DA. The smaller number of items may explain the slightly lower internal consistency. Similar Alpha

Coefficients for the DA have been reported in other studies, but internal consistency has not been reported in prior work for either the DVSI or the K-SID. The K-SID was far less internally consistent (Standardized Alpha = .05) and also the shortest of the three instruments at only 10 items.

The DV-MOSAIC IQ Score calculates the number of missing answers, with the premise that over a certain percentage of missing responses calls into question the accuracy of the results, showed adequate data for assessments from most respondents. Only 2.7% of the participants overall had a lower IQ Score than is considered acceptable by de Becker and associates. The highest percentages of unacceptably low IQ scores were found among Hispanic/Latina respondents (4.2%) and those who spoke Spanish (3.6%).

Concurrent Validity.

All of the instruments and DV-MOSAIC had evidence of concurrent validity in terms of significant correlation with frequency and severity of physical abuse at baseline as measured by the CTS2, with the DA having the strongest correlation (.459) and the K-SID the weakest (.134) with the DVSI (.37) and DV-MOSAIC (.355) in between. None of these correlations is strong, but it can be argued that they should not be: if these threat and risk assessment strategies were redundant with a measure of severity and frequency of current abuse, a measure of past abuse could be used instead of a risk or threat assessment. In addition, it can be posited that concurrent construct validity is not as important in evaluation risk assessment methods as it is with conventional instruments (Hilton et al., 2004). Even so we correlated the mean score of the risk assessment methods with the woman's prediction of the likelihood of physical abuse and severe harm in two separate items as another measure of concurrent validity. Except for the K-SID, all the instruments had significant but weak correlations in the expected direction, with the DV-MOSAIC (.18 and .15 for physical abuse & severe harm, respectively) and the DA (.15 and .18) having the strongest correlations and the DVSI somewhat weaker (.15 and .12). These correlations were far weaker than expected, given the findings from previous studies of the accuracy of victims' prediction of risk (Gondolf & Heckert, 2003; Goodman, Dutton & Bennett, 2000; Weisz, Tolman, & Saunders, 2000). Unlike these studies, we asked two separate questions on risk of re-assault and risk of serious harm, but the questions were derived from those previous studies.

Assessing Predictive Ability – Bivariate Correlations

The DA, DV-MOSAIC and DVSI correlated significantly with severity of abuse at follow-up, although these correlations were weak. The K-SID did not. Removing the 27 women never exposed to the possibility of re-assault during the follow-up period increased the correlations to .38 for the DA, .21 for the DV-MOSAIC and .15 for the DVSI. All three instruments had even lower correlations with frequency and severity of physical assault (and frequency and severity of sexual assault), with the DV-MOSAIC correlations failing to reach significance. However, the frequency and severity of future assault (as a continuous variable) is not what these instruments and DV-MOSAIC are designed to assess (but rather whether or not severe assault will occur), so the latter findings are interesting but probably not a useful standard to hold the risk assessment methods to.

The mean scores of the DA, DV-MOSAIC and the DVSI all correlated significantly with the 8 levels of severity of abuse that we proposed as one of our major outcome variables. In addition, the DA levels of danger, the highest (8-10) ratings versus scores lower than 8 based on the DV-MOSAIC

domains of inquiry and the DVSI low and high risk levels all showed a pattern of correctly classifying participants into the levels of abuse. The K-SID did not show either of these patterns. One issue with these analyses is the validity of the severity of abuse categories developed by the research team. Although there were significant correlations with the CTS2, the WEB scale and HARASS, those correlations were not strong. The threats and stalking level was associated with higher levels of risk than low physical abuse on all of the instruments/method. Placing stalking and threats at a higher level of severity of outcomes than low level physical abuse may actually be a more accurate placement given the associations of stalking and threats to kill with intimate partner femicide (Campbell et al., 2003).

Another issue is that, although our measures of severe violence included potentially lethal acts, we did not attempt to find actual or attempted homicides among our cases, and our study cannot really address the prediction of those outcomes. Interestingly, however, there were some risk factors that are associated with lethality in the literature but were not associated with risk of reassault in this study. Specifically, the abuser having access to a gun and prior threats with a weapon are predictive of lethality (Campbell et al., 2003), but were not predictive of reassault. That finding is simple to interpret: use of a weapon facilitates murder, but is not necessary to assault. In addition, suicidality and severe depression are predictive of lethality, particularly of homicide-suicide, but are not associated with the risk of reassault or severity.

Reassault and Women's Perceptions and Actions

Almost one third (30.9%) of our participants had experienced some sort of physical or sexual re-assault at Time 2, including 86 women (11%) who had experienced a severe, potentially lethal act. This percentage of re-assault is within the range of what has been reported in other research (Dutton, 1995; Block, 1999; Saunders, 1995; Ford & Regoli, 1992; Davis, Taylor, & Maxwell, 2000). Breaking out the different levels of severity of re-assault and other forms of abuse provides new and potentially extremely useful information. If we remove the 27 women for whom there was no chance of re-victimization because the perpetrator was in jail the entire time or otherwise totally unable to contact her, the proportion of those re-assaulted is only slightly higher (32.6%) or almost exactly one third. We also included verbal and psychological abuse and harassment in our outcome measures and found that only 16% of the women reported no repeat abuse of any kind at time two.

Women's perceptions of risk according to the questions we asked about likelihood of re-assault or serious harm were not significantly correlated with severity of re-assault in bivariate correlations but, as has been reported in previous research (Gondolf & Heckert, 2003; Goodman, Dutton & Bennett, 2000; Weisz, Tolman, & Saunders, 2000), victims' predictions were significantly associated with re-assault as a dichotomous outcome, with a risk ratio of 1.8 for any re-assault. Collapsing the two highest and two lowest levels of danger on the DA, there was a risk ratio of 2.5 for any re-assault and 4.6 for severe assault. The DVSI significantly predicted risk of severe re-assault only (risk ratio of 1.8 compared to 1.9 for victim perception of risk of severe re-assault).

In addition to the victims being impressive predictors of their own risk, it is important to note the remarkable extent and array of the actions these women took to escape future assault and abuse from these perpetrators. Women in the highest-risk categories of the DA and the DVSI and at the highest ratings of DV-MOSAIC were more likely to take actions to escape their abusive ex-partners

or otherwise decrease their risk. They were significantly more likely to go someplace where the abusers could not find them (e.g. 41.3% in the extreme danger category of the DA vs. 4.3% in the variable danger category; 44.4% of those at the highest ratings on DV-MOSAIC vs. 23% at the lower ratings). Women in the highest-risk categories for the DA and the DVSI and the highest ratings for the DV-MOSAIC were more likely than other victims to go to a shelter for battered women (e.g. 6.8% at low risk versus 15.6% of those at high risk on the DVSI). The likelihood that the victim had no voluntary contact with her abusive ex-partner increased to approximately 50% at higher levels of risk at baseline DA and DVSI and the K-SID as well as at the highest ratings on DV-MOSAIC. Women in high-risk and ratings categories at baseline also tended to be more likely than other women during the follow-up period to receive counseling, do safety planning, change the locks on their doors, or obtain a weapon for protection, usually mace or pepper spray. The system also responded to these cases that were at the highest levels of risk on all four baseline assessments with about 25% of the highest risk categories going to jail compared to about 10% of the lower risk level perpetrators. Thus, women's actions suggested that they were predicting risk of re-assault in concordance with the risk assessment instruments and strategies we were evaluating, and taking effective steps to reduce the risk of assault. These actions were also correlated with victims' own perception of risk that their partner or ex-partner would harm them.

To the extent that victims were successful in escaping or diminishing their chance of re-assault, the actions of the participants and system interventions may have resulted in an understatement of the ability of the instruments and method to discern risk for future assault because the most dangerous abusers and circumstances are likely to prompt greater protective efforts. That protective actions were more likely in higher risk cases and that they were often effective is by and large a hopeful sign for the field. It is, however, a limitation of the study. We tried to take these actions into account by eliminating the 27 women from our analyses who were never at risk for re-assault and, for the rest of the sample, taking into account time at risk and other protective actions taken by the victim or the system. The first adjustment is presented in the bivariate analyses and both are adjusted for in the multivariate analyses.

Sensitivity, Specificity and Positive Predictive Power

Predictive instruments are often evaluated on their sensitivity, specificity and positive predictive power based on their ability to identify the largest proportion of true positives possible (sensitivity) while not including false positives and at the same time correctly identifying as many true negatives as possible as being at low risk. An instrument with high sensitivity casts the kind of wide net that is needed to be sure there are few false negatives (victim is declared at low risk but is re-assaulted) while high specificity insures that there are few false positives (few perpetratorsare falsely declared to be high risk and deprived of liberty unnecessarily or on false premises). There was a wide range of sensitivity on our three instruments and one method with most risk levels of most of the methods having the usual tradeoff of low specificity with high sensitivity and vice versa. The best balance was the victims' perceptions of risk (.665 sensitivity and .532 specificity) and the high and extreme levels of risk for the DA (.871 and .492 sensitivity and .335 and .651 specificity respectively). At the highest ratings (8, 9, 10), the DV-MOSAIC had acceptable specificity (.653) but relatively low sensitivity (.371) while the DVSI had better sensitivity (.554) but lower specificity (.457). It is perhaps appropriate for the DVSI to have higher sensitivity but lower specificity since the DVSI was designed to initially identify DV cases that might be at high risk with the idea that the SARA would then be used with those cases for a more precise assessment of the level of risk.

Conversely, the DV-MOSAIC was mainly designed for the criminal justice system where specificity is most important. Even at the highest levels or ratings of risk, and with either victim reported reassault or arrests, all of the approaches had a fairly high proportion of false positives. False positives (identifying as high risk cases when asssault did not occur during follow up), were approximately 30% for the DA and DV-MOSAIC, and about 25% for the K-SID. These rates are of concern if they are used to deny liberty to abusers unnecessarily. In addition, there was a 16-33% false negative rate depending on the method. These cases obviously are of serious concern because of victim safety. Yet criminal justice decision making relevant to cases of intimate partner violence in the absence of standardized risk assessment results in many false positives and false negatives. The performance of the tested risk assessment methods were not compared with that of judges or probation officers, so we cannot say whether the predictions generated by these methods are more or less accurate than criminal justice officials.

Positive Predictive Value (PPV) considers sensitivity and specificity (predictive validity) of the instrument but also the probability of the event being predicted. The PPV of the DA, DV-MOSAIC, DVSI and K-SID were all within the same range, approximately .10 apart, with the DA Extreme Danger having the highest PPV (.438), the victim's perception of high risk was next (.391) and the K-SID high risk level the next (.366). The PPV of the DVSI and DV-MOSAIC high risk levels were very close at .345 and .317 respectively.

All of the instrument/method-defined risk categories had better sensitivity in predicting severe/potentially lethal abuse than in predicting any physical assault. The victim perception of risk (using the perception of risk of serious harm item) was very much the same and again quite impressive (sensitivity .63, specificity .563). However, sensitivity on the DA was quite a bit higher (> .90) at the lower levels of risk (Increased and Severe Danger). This pattern might be different for the victims' perception had we divided the continuous scale into four categories instead of three. The DVSI and K-SID at moderate risk were approximately the same in sensitivity but all 3 were lower in specificity. The highest ratings on the DV-MOSAIC and the Extreme level of Danger on the DA had higher specifictiy than the victim perception high category (6-10). In other words, victims' ratings of 6-10 led to a higher proportion of false positive predictions than the highest level of either DV-MOSAIC or the DA. It is also interesting that the K-SID best predicted rearrest, exactly as it was designed to do.

ROC Analysis

Receiver Operator Curves (ROC) are a means of combining sensitivity and specificity information without depending on the cut points of the instrument or risk assessment method. They are also fairly straightforward to interpret since one is looking for an area under the curve significantly more than .50 in prediction of a future event. The ROC analysis that is probably most reasonable to consider is that which excludes the 27 women who were not exposed to the perpetrator. Here we find that for any abuse, only the DA and victim's perception of risk are significantly better than chance and the DA improves upon the victim's perception of risk but only marginally (.635 vs. 599). For severe re-assault, the DVSI also significantly improves on chance but is similar to the victims' perception of risk of serious harm. The DA area under the curve for severe assault is .67. When we added predictive probabilities from multivariate analysis to take into account potential confounds such as time at risk, all of the ROC curves of all of the instruments and method become significantly greater than chance with the areas under the curves ranging from .69 (DA) to .58 (DV-MOSAIC predicting any assault). However, the specific effects of the risk assessment scores in the

regression models that produced the probabilities for the ROC analyses were statistically significant only for the DA (both any abuse and severe abuse) and the DVSI for severe abuse. All of the risk assessment methods had modestly (no more than .02) larger areas under the curve for severe than for any re-assault.

Including the criminal justice data reduced the predictive accuracy of all the methods and the victims' assessments. Arrest data were available for all the participants, even those not reached at follow-up which increased the sample size and therefore increased the power to detect significance. But for those not reached at follow-up, only arrest data were available, and arrest data drastically underestimated the reassault rate compared to victim reports. In fact, only 18% of the abusers were arrested for a violent crime, and we know from those victims we reached at follow-up that this rate fails to capture the rates of reassault reported by victims. Therefore, tests using criminal justice as the only measure of reassault for 40% of the participants inflated the false positive rate and decreased specificity. That is, re-assault was predicted by the method but was not represented in the arrest data and therefore artificially lowered the area under the ROC curve for all methods.

Overall, the DA and victims' estimates were consistently better than .50, with the DA performing somewhat better than victims' estimates. When controlling for the protective actions taken, all the methods achieved significance in predicting any assault and severe assault. All the approaches predicted severe assault better than they predicted any abuse, especially the DV-MOSAIC. Controlling for protective actions improved the predictive ability of all the methods by decreasing the false positive rate. That is, when a method predicts risk but an action is taken that reduces or eliminates the possibility of reassault, the method appears to be overestimating risk (increases false positives). By taking into account or controlling for such protective actions, the ROC curves better reflect the real predictive accuracy of the methods. The accuracy of victims' assessments also improved when controlling for protective actions, but they improved the least of all the methods. The reason for this may be that victims were taking into account their protective action plans when they made their assessment.

Multivariate Models of Levels of Risk

When the levels of risk were tested in multivariate models, the overall models were significant for all except the K-SID. The DA again was the most predictive but only significant (< .05) at the highest two levels (high and extreme) and again was most predictive for severe re-assault (aOR = 11.0 and 17.2). The DA was also strongly predictive of stalking and threats at the two highest levels. The DVSI and DV-MOSAIC were also significantly predictive of severe re-assault and stalking. The K-SID levels of risk were not significantly predictive of severe re-assault but the higher two levels were predictive of minor/moderate abuse (aOR = 4.9 and 3.2) and at the highest level was significantly predictive of stalking and threats (aOR = 1.8).

Predicting Risk of Sexual Re-Assault

Generally, only the DA was successful at predicting intimate partner sexual re-assault, probably partly because there is a question on prior intimate partner sexual assault because the forced sex aspect of IPV has been part of the DA author's conceptualization. Given that 41% of our respondents reported having been physically forced into sex by their partner at baseline and an additional 20% otherwise coerced (without physical force or assault at the time of the incident but by threat of force or other coercive tactics), intimate partner sexual assault is indeed a concern for

these women. However, the other risk assessment instruments and method were not designed with sexual assault in mind and should not be evaluated on their ability to predict this particular form of domestic violence. Given the frequency of sexual abuse in intimate partner violence, particularly among these relatively severely abused women, it is an important factor to consider in future development of these types of instruments and methods and in the outcomes used in future research.

Summary

The participants were severely abused women at Time 1, with 82% having experienced severe abuse and all but 6% having experienced some form of physical assault by their partner or ex-partner. Approximately one third were re-assaulted by the end of the four to 24 month period when they were re-interviewed. They continued to be severely abused with 11% having experienced a severe, potentially lethal act. Although there were limitations to the study, valuable information about the three risk assessment instruments and DV-MOSAIC was gained. By most analytic strategies, the Danger Assessment had the strongest psychometric properties, including the predictive statistics. By most analytic strategies, DVSI and DV-MOSAIC also had significant associations with future re-assault. The K-SID was least strong under most of the analytic strategies; however, it did best at predicting rearrest using the criminal justice data. Some of the differences in results among them could be attributed to the different purposes and settings where these three instruments and the one threat assessment method were developed to be used. The K-SID was designed to be used in criminal justice settings, to assist in probation and parole (post-incarceration) decisions. The goal is to predict rearrests. The DA is the only instrument that was meant to be an interview of victims which is the way the study was implemented. The K-SID and DVSI and DV-MOSAIC are meant to be filled out by criminal justice personnel primarily using information from criminal justice records, although DV-MOSAIC is also meant to be based on information from the victim. In addition, the DA and DV-MOSAIC have more severe abuse or potential lethality as the potential outcome being assessed, while the DVSI and K-SID are meant to predict risk of any re-assault. In fact, that the DVSI and K-SID did better at predicting rearrest using criminal justice data supports their use for those purposes and in those settings.

The differences in purpose versus implementation of the study may help explain why none of the instruments or method was particularly impressive in predicting re-assault. By most analytic strategies, women's perception of risk did better than the other assessment methods or almost as well as the DA, the most predictive of those tested. But even the women's prediction left much of the re-assault unanticipated. Thus, there is much research that still needs to be done in the field. This study joins only about 8 other published studies that we could find that have been done to address the need for assessment of risk in the field of intimate partner violence. We have a long way to go before we can match the 95 studies in the recent Hanson and Morton-Bourgon (2004) meta-analysis of prediction of sexual assault.

We found that the approaches to risk assessment we tested were high on sensitivity (.80 -.90 for the DA and DV-MOSAIC with cut-offs at the second lowest levels of risk), in other words they correctly classified most of the women that were indeed reassaulted as being at somewhat elevated risk. The great majority of women with the lower risk scores or ratings were not subsequently re-assaulted (over 80% on the DA and DV-MOSAIC; 67–72% on the DVSI and K-SID). These high "true negative" and low "false negative" rates (i.e., high sensitivity) bode well for

victim safety, although certainly the 16-33% false negatives depending on method, the women who were predicted to be at low risk yet experienced violence, even severe violence, are always a concern.

At these low levels of risk, the methods were also very low in specificity, a more tolerable form of error. At the highest levels of risk, specificity was higher, as one would hope. But even at the highest levels or ratings of risk, with either victim reported reassaults or arrests as the outcome measure, all of the approaches had a fairly high proportion of false positives (approximately 30% for the DA and DV-MOSAIC, about 25% for the K-SID). These low specificity figures (high false positives) are more of a concern for offender rights than victim safety, although they also have an impact on victims. If these risk assessment approaches are used as the basis for decisions, about one third of women being assessed might be unduly frightened and make major changes in their lives that may not be necessary, scarce resources will be expended on low risk cases, credibility will be lost, and partners might be treated with unduly harsh penalties.

Risk assessment instruments or methods should not be the only factors considered in making decisions about victim safety or offender sanctions, especially at our current state of the science. Rather, they are meant to be one source of information among many others. When confounding variables were controlled in the ROC analysis, all four of the instruments performed significantly better than .50 in predicting any or severe physical or sexual abuse during the follow-up period. Yet, this is not good enough. Risk assessment instruments need to perform significantly better than expert judgment, the view of experienced victim advocates, law enforcement officers, probation officers, or other practitioners, or they are not worth the time and effort they take. This study shows that systematic risk assessment approaches can be better predictors than the victims, but does not address whether they are better than experienced practitioners. However, in the field of sexual assault and mental health, instruments have been found to be significantly better than expert judgment, and a combination of instrument or systematic formal method and expert judgment is thought to be the best approach (Pinard & Pagani, 2000; Hanson and Morton-Bourgon, 2004). Kropp (2004) calls this combined approach "structured professional judgment," which is very similar to the threat assessment approach described by de Becker to be used with DV MOSAIC. Before a particular approach can be recommended unequivocally for use in the domestic violence realm, it must be tested in the field, comparing the predictive accuracy of the instrument to or in combination with expert judgment.

Additional analysis is also needed on the current data to examine individual risk factors more closely. Across the four instruments and non-redundant items, over 100 risk factors were included in the interviews. Additional item analysis may shed light on which individual risk factors are most predictive and which might be combined into new instruments useful in different settings and for different purposes.

We note that the women in this study took significant steps to protect themselves from further abuse. In fact, most participants were recruited from sites where they were already receiving assistance or taking action. We do not know what the effect would be on victims' ratings of risk had they been given their scores on the risk assessments. Future research on the role of risk assessments in increasing victims' self-protective actions is particularly needed.

Until further field research is completed, we recommend that practitioners assess risk systematically, with one of the methods used in this research or another method with some evidence of validity, and also carefully explore the victim's perception of risk (see Practitioner Summary). This information should be combined with all other aspects of the case, the more information the better, and the practitioner's own expert judgment, as the instruments and method have a level of fallibility that means they should never be taken as definitive.

REFERENCES

Anderson, J.C., & Gerbing, W. (1988). Structural equation modeling in practice: A review and recommended two-step program. Psychological Bulletin, 103, 411-423.

Arbuckle, J., Olson, L., Howard, M., Brillman, J., Ancti, C., & Sklar, D. (1996). Safe at home? Domestic violence and other homicides among women in Mexico. Annals of Emergency Medicine, 27(2), 210-215.

Bachman, R., & Saltzman, L. E. (1995). Violence against women: Estimates from the redesigned survey. Washington, DC: US Department of Justice.

Bailey, J. E., Kellermann, A. L., Somes, G. W., Banton, J. G., Rivara, F. P., & Rushford, N. P. (1997). Risk factors for violent death of women in the home. Archives of Internal Medicine, 157, 777-782.

Berk, R. A., Berk, S., Loseke, D. R., & Rauma, D. (1983). Mutual combat and other family violence myths. In D. Finkelhor, R. J. Gelles, G. T. Hotaling, & M. A. Straus (Eds.), The dark side of families (pp. 197-212). Beverly Hills, CA: Sage.

Block, C. R., & Christakos, A. (1995). Intimate partner homicide in Chicago over 29 years. Crime & Delinquency, 41 (4),406-526.

Bollen, K.A. (1989). Structural equations with latent variables. New York: Wiley.

Borenstein, M., Rothstein, H. & Cohen, J. (1997). Power and precision: A computer program for statistical power analysis and confidence intervals. Teaneck, NJ: Bistat.

Bowman, P. J. (1993). The impact of economic marginality among African American husbands and fathers. In H. McAdoo (Ed.), Family ethnicity: Strength in diversity (pp. 120-140).

Brock, K. (2003). When men murder women: An analysis of 2001 homicide data. Washington, DC: Violence Policy Center Publications.

Browne, A. (1987). Battered women who kill. New York: Free Press.

Browne, A., Williams, K. R., & Dutton, D. C. (1998). Homicide between intimate partners. In M. D. Smith & M. Zahn (Eds.), Homicide: a sourcebook of social research (pp. 149-164). Thousand Oaks, CA: Sage.

Buteau, J., Lesage, A. D., & Kiely, M. C. (1993). Homicide followed by suicide: A Quebec case series, 1988-1990. Canadian Journal of Psychiatry, 38(8), 552-556.

Buzawa & Buzawa (1996). Do arrests and restraining orders work? Thousand Oaks, CA: Sage Publications.

Campbell, D., Campbell, J., Parker, B., King, C. & Ryan, J. (1994). The reliability and

factor structure of the Index of Spouse Abuse with African American women. Violence and Victims, 9(3), 259-274.

Campbell, J. C. (1995). *Assessing dangerousness*, Newbury Park: Sage.

Campbell, J. C. (1981). Misogyny and homicide of women. Advances in Nursing Science, 3(2), 67-85.

Campbell, J. C. (1986). Nursing assessment of risk of homicide for battered women. Advances in Nursing Science, 8(4), 36-51.

Campbell, J. C. (1992). "If I can't have you, no one can": Power and control in homicide of female partners. In J. Radford & D. E. H. Russell (Eds.), Femicide: The Politics of Woman Killing (pp. 99-113). New York: Twayne.

Campbell, J. C. (1994). Domestic homicide: Risk assessment and professional duty to warn. Maryland Medical Journal, 43(10), 885-889.

Campbell, J. C. (1995). Prediction of homicide of and by battered women. In J. C. Campbell (Ed.), Assessing the risk of dangerousness: Potential for further violence of sexual offenders. batterers. and child abusers (pp. 96-113). Newbury Park, CA: Sage.

Campbell, J. C. (2002). Safety planning based on lethality assessment for partners of batterers in treatment. Journal of Aggression, Maltreatment, and Trauma 5 (2), 129-143.

Campbell, J. C. (1996). Risk factors for homicide in violent intimate relationships. Washington, DC: National Institute for Health Funded Research. #5R01 DA 11156-02.

Campbell, J. C., Rose, L. E., Kub, J., & Nedd D. (1998). Voices of strength and resistance: A contextual and longitudinal analysis of women's responses to battering. *Journal of Interpersonal Violence* 13, 743-762.

Campbell, J. C., Sharps, P., & Glass, N. (2000). Risk Assessment for Intimate Partner Violence. In G.F. Pinard & L. Pagani (Eds.), Clinical Assessment of Dangerousness: Empirical Contributions. New York: Cambridge University Press.

Campbell, J. C. & Soeken, K. (1999). Forced sex and intimate partner violence: Effects on women's health. Violence Against Women 5 (9),1017-1035.

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

Campbell, J. & Soeken, K. (1999). Women's responses to battering over time: An analysis of change. Journal of Interpersonal Violence, 14(1), 21-40.

Campbell, J., Torres, S., Ryan, J., King, C., Campbell, D., Stalling, R. & Fuchs, S. (1999). Physical and Nonphysical abuse and other risk factors for low birthweight among term and preterm babies: A multiethnic case control study. American Journal of Epidemiology, 150(7) 714-726.

Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C. R., Campbell, D. W., Curry, M. A. Gary, F. A., Glass, N. E., McFarlane, J., Sachs, C., J., Sharps, P. W., Ulrich, Y., Wilt, S. A., Manganello, J., Xu, X., Schollenberger, J., Frye, V., & Laughon, K. (2003). Risk factors for femicide in abusive relationships: Results from a mulitsite case control study. American.Journal.of

Public Health 93 (7), 1089-1097.

Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C. R., Campbell, D. W., Curry, M. A. Gary, F. A., Glass, N. E., McFarlane, J., Sachs, C., J., Sharps, P. W., Ulrich, Y., Wilt, S. A. (2003). Assessing risk factors for intimate partner homicide. National Institute of Justice Journal, 250, 14-19.

Clayton, D. & Hills, M. (1993). Statistical Methods in Epidemiology. Oxford: Oxford University Press.

Crawford, M., & Gartner, R. (1992). Woman killing: Intimate femicide in Ontario: 1974-1990. Ontario, Canada: Women We Honor Action Committee.

Crowell, N. A. & Burgess, A. W. (1996). Understanding violence against women. Washington, DC: National Research Council.

Daly, M., Wiseman, K. A., & Wilson, M. (1997). Women with children sired by previous partners incur excess risk of uxoricide. Homicide Studies, 1(1), 61-71.

Davis, R. C., Taylor, B. G., & Maxwell, C. D. (2000). Does batterer treatment reduce violence? A randomized experiment in Brooklyn. Final report to the National institute of Justice. Washington, D.C.: US Department of Justice.

Davis, R. C., & Taylor, B. G. (1997). A proactive response to family violence: The results of a randomized experiment. Criminology, 35, 303-333.

Dawson, R., & Gartner, R. (1998). Differences in the characteristics of intimate femicides: The role of relationship state and relationship status. Homicide Studies, 2, 378-399.

Dearwater, S., Coben, J., Nah, G., Campbell, J., McLoughlin, E. & Glass, N. (1998). Prevalence of domestic violence in women treated at community hospital emergency departments. Journal of the American Medical Association, 280(5), 433-438.

De Becker, G. (1997). The gift of fear. Boston: Little, Brown & Co.

De Becker, G. & Associates. (2000). *Domestic Violence Method (DV MOSAIC).* http://www.mosaicsystem.com/dv.htm

DeKeseredy, W. S. & Schwartz, M. D. (1998). Measuring the extent of woman abuse in intimate heterosexual relationships: A critique of the Conflict Tactics Scale. VAWnet: http://www.vaw.umn.edu/Vawnet/ctscrit.html.

Dobash, R. E., Dobash, R. P., Cavanagh, K. & Lewis. R. (1998). Women's perceptions of men's violence. Violence Against Women 4 (4):382-414.

Dugan, L., Nagin, D., & Rosenfeld, R. Do domestic violence services save lives? *National Institute of Justice Journal* (250):20-25, 2003.

Dutton, D. G. (1995). The domestic assault of women. Vancouver: UBC Press.

Dutton, D. G. (1995b). Trauma symptoms and PTSD-like profiles in perpetrators of intimate abuse. Journal of Traumatic Stress, 8(2), 299-316.

Dutton, D. G., Bodnarchuk, R., Kropp, R., Hart, S. D. & Ogloff, J. P. (1997). Client personality disorders affecting wife assault post-treatment recidivism. Violence & Victims, 12(1):37-50.

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

Dutton, D. G., Landolt, M. A., Starzomski, A., Bodnarchuk, M. (2001). Validation of the PAS in diverse male populations. Journal of Family Violence, 16(1), 69-73.

Fagan, J., Stewart, D. E., & Hansen, K. (1983). Violent men or violent husbands? Background factors and situational correlates. In R. J. Gelles, G. Hotaling, M. A. Straus, & D. Finkelhor (Eds.), The dark side of families (pp. 49-68). Beverly Hills: Sage.

Ferraro, K. J., & Johnson, J. M. (1983). How women experience battering: The process of victimization. Social Problems, 30, 325-339.

Gartner, R. (1990). The victims of homicide: A temporal and cross national comparison. American Sociological Review, 55, 92-106.

Gelles, R. (1998, October). Lethality and risk assessment for family violence cases. Paper presented at the 4th International Conference on Children Exposed to Family Violence, San Diego, CA.

Gondolf, E. W. (1997). Patterns of reassault in batterer programs. Violence and Victims 12 (4), 373-387.

Gondolf, E. W. (2000). How batterer program participants avoid reassault. Violence Against Women 6 (11), 1204-1222.

Gondolf, E. W., & Heckert, D. A. (2003). Determinants of women's perceptions of risk in battering relationships. Violence & Victims 18 (4):371-386, 2003.

Goodman, L., 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 10, 63-74.

Grayson, R., O'Sullivan, C., Fish, C., & Frank, P. (1999). Test Site for Community Oriented Policing Domestic Violence Initiatives in Underserved Communities: African American and Haitian Communities in Spring Valley. Village of Spring Valley, NY: COPS project.

Grisso, J. A., Schwarz, D. F., Hirschinger, N., Sammel, M., Brensinger, C., Santanna, J., Lowe, R. A., Anderson, E., Shaw, L. M., Bethel, C. A., & Teeple, L. (1999). Violent injuries among women in an urban area. New England Journal of Medicine,341 1899-1905.

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

Grisso, J.A., Schwarz, D., Hirschinger, N., Sammel, M., Brensinger, C., Santanna, J., Lowe, RA., Anderson, E., Shaw, L.M., Courtney A.B., Teeple, L. (1999) Violent Injuries Among Women in an Urban Area. New England Journal of Medicine, 341:1899-1905

Hall-Smith, P., Earp, JA, & DeVellis, R. (1995). Measuring battering: Development of the Women's Experience with Battering (WEB) Scale. Women's Health: Research on Gender, Behavior, and Policy, 1, 273-288.

Hall-Smith, P., Moracco, K. E., & Butts, J. (1998). Partner homicide in context. Homicide Studies, 2(4), 400-421.

Hall-Smith, P., Smith, H. B., Earp, JA L. (1999). Beyond the measurement trap: A reconstructed conceptualization and measurement of women battering. Psychology of Women Quarterly ,23, 177-193.

Hall-Smith, P., Tessaro, I., & Earp, JA (1995). Women's experiences with battering: A conceptualization from qualitative research. Women's Health Issues, 5, 173-193.

Hanson, R. K., & Morton-Bourgon, K. (2004). Predictors of sexual recidivism: An updated meta-analysis. (User Report 2004-02). Ottawa: Public Safety and Emergency Preparedness Canada. http://www.psepc.gc.ca/publications/corrections/pdf/200402_e.pdf

Hanson, R. K., Cadsky, 0.,. Harris, A. & Lalonde, C. (1997). Correlates of battering among 997 men: family history, adjustment, and attitudinal differences. Violence and Victims, 12(3), 191-208.

Hare, R. D. (1991). Manual for the Hare Psychopath Checklist-Revised PCL-R. Toronto: Multi-Health Systems.

Hart, B. (1988). Beyond the "duty to warn": A therapist's "duty to protect" battered women and children. In K. Yllo & M. Bogard (Eds.), Feminist perspectives on wife abuse (pp. 234-248). Newbury Park: Sage.

Healey, K., Smith, C., & O'Sullivan, C. (1998). Batterer Intervention: Program Approaches and Criminal Justice Strategies. Washington, D.C.: National Institute of Justice.

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 (7), 778-800.

Holtzworth-Munroe, A., & Stuart, G. L. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin* 116 (3):476-497.

Holtzworth-Munroe, A., Meehan, J. C., Herron, K., Rehman, U., & Stuart, G. L. (2000). Testing the Holtzworth-Munroe and Stuart (1994) Batterer Typology. *Journal of Consulting & Clinical Psychology* 68 (6):1000-1019.

Hotaling, GB., & Sugarman, D. B. (1990). A risk marker analysis of assaulted wives. Journal of Family Violence 5,1-13.

Hoyle, R.H. (Ed.) (1995) Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage.

Howard, M. (1986). Husband-wife homicide: An essay from a family law perspective. Law and Contemporary Problems, 49(1), 63-88.

Humphreys, J. C. & Campbell, J. C. (2004). *Family violence in nursing practice*, Philadelphia:Lippincott, Williams & Wilkins.

Jackson, S., Feder, L., Forde, D.R., Davis, R. C., Maxwell, C. D. & Taylor, B. G. (2003). Batterer Intervention Programs: Where Do We Go From Here? NIJ Special Report. Washington, D.C.: USDOJ.

Jacobson, N. S., Gottman, J. M., Gortner, E., Berns, S., & Shortt, J. W. (1996). Psychological factors in the longitudinal course of battering: When do the couples split up? When does the abuse decrease? Violence and Victims 11 (4),371-392.

Jaynes, G. D., & Williams, R. M. (1989). A common destiny: Blacks and American society. Washington, DC: National Academy Press.

Joreskog, K.G. & D. Sorbom. (1976). LISREL III Estimation of linear structural equation systems by maximum likelihood methods. Chicago: National Education Resources.

Jurik, N. C. & Winn, R. (1990). Gender and homicide: A comparison of men and women who kill. Violence & Victims, 5, 227-242.

Kellerman, A. L., Rivara, F. P., & Rushforth, N. B. (1993). Gun ownership as a risk factor for homicide in the home. New England Journal of Medicine, 329, 1084-1091.

Klein, E., Campbell, J, Soler, E., & Ghez, M. (1997). Ending domestic violence: Changing public perceptions. Newbury Park: Sage.

Koziol-McLain, J., Webster, D., McFarlane, J., Block, C. R., Curry, M. A., Ulrich, Y., Wilt, S., Glass, N. Campbell, J. C. Risk factors for femicide-suicide in abusive relationships: Results from a multi-site case control study. Submitted to *JAMA.*

Kropp, P. R. (1998, October). Risk assessment and monitoring in domestic violence cases: Recent research and practice. Paper presented at the 4th International Conference on Children Exposed to Family Violence, San Diego, CA.

Kropp, R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment Guide (SARA): Reliability and validity in adult male offenders. Law and Human Behavior, 24, 101-118.

Kropp, P. R. (2004). Some questions regarding spousal assault risk assessment. Violence Against Women, 10 (6), 676-697.

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

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (2000). The spousal assault risk assessment guide (SARA): Reliability and validity in adult male offenders. *Law and Human Behavior,* 24: 101-118.

Kyriacou, D. N., Anglin, D., Taliaferro, E, Stone, S., Tubb, T., Linden, J. A., Muelleman, R., Barton, E., & Kraus, J. F. (1999). Risk factors for injury to women from domestic violence against women [see comments]. New England Journal of Medicine 341 1892-1898.

Laming, D, (1973). Mathematical Psychology. London: Academic Press.

Langford, D, R. (1996). Predicting unpredictability: A model of women's processes of predicting battering men's violence, Scholarly Inquiry for Nursing Practice10 (4),371-385.

Langford, L., Isaac, N. E., Kabat, S. (1998). Homicides related to intimate partner violence in Massachusetts. *Homicide Studies* 2 (4):353-377, 1998.

Little, R.JA.,& Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: Wiley, Litwack, T. R., Kirschner, S. M., & Wack, R. C. (1993). The assessment of dangerousness and predictions of violence: Recent research and future prospects. Psychiatric Quarterlv,64 245-273.

Litwack, T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A. Hess & I. Weiner (eds.), The handbook of forensic psychology (p. 171-217). New York: John Wiley & Sons.

Lyon, E, (1998, May). The Revised K-SID: Analysis of Reliability and Relationships to New Arrests After One Year (An Interim Report). The Office of Policy and Management State of Connecticut.

Marin, G. & Marin, B. V. (1991). Research with Hispanic populations, Newbury Park: Sage Publications.

Maxwell, C.D., Garner, J. H. & Fagan, J. A. (2001). The Effects of Arrest on IPV: New Evidence from the Spouse Assault Replication Program. NIJ *Research in Brief.* Washington, D.C.: USDOJ.

McFarlane, J. Campbell, J. C., Wilt, S., Sachs, C., Ulrich, Y., & Xu, X. (1999). Stalking and intimate partner femicide. Homicide Studies3 (4).

McFarlane, J., Campbell, J. C., Sharps, P. W., & Watson, K. (2002). Abuse during pregnancy and femicide: urgent implications for women's health. *Obstet.Gynecol.* 100 (1):27-36.

McFarlane, J., Parker, B., Soeken, K., & Bullock, L. (1992). Assessing for abuse during pregnancy: Severity and frequency of injuries and associated entry into prenatal care. Journal of the American Medical Association,267, 2370-2372.

McFarlane, J., Parker, B., Soeken, K., Silva, C., & Reel, S. (1997). Safety behaviors of abused women after an intervention during pregnancy. JOGNN 27 (1):64-69.

McFarlane, J., Soeken, K., Campbell, J. C., Parker, B., Reel, S. & Silva, C. (1998), Severity of abuse to pregnant women and associated gun access of the perpetrator. Public Health Nursing 15 (3):201-206.

Mercy, J. A., & Saltzman, L. E. (1989). Fatal violence among spouses in the United States 1976-85. American Journal of Public Health, 79, 595-599.

Moffitt, T., & Casper, A. (1999, July). NIJ Research in Brief: Findings about partner violence from the Dunedin Multidisciplinary Health and Development Study. Washington, D.C.: US Department of Justice.

Monahan, J. (1996). Violence prediction: The past twenty years and the next twenty years. Criminal Justice and Behavior, 23(1), 107-120.

Moracco, K. E., Runyan, C. W., & Butts, J. (1998). Femicide in North Carolina. Homicide Studies, 2, 422-446.

Morton, E., Runyan, C. W., Moracco, K. E., & Butts, J. (1998). Partner homicide victims: a population based study in North Carolina, 1988-1992. Violence & Victims,13(2) , 91-106.

Moss, V., Pitula, C., Campbell, J. & Halstead, L. (1997). The experience of terminating an abusive relationship from an Anglo and African-American perspective: A qualitative study. Issues in Mental Health Nursing,18, 433-54.

Mulvey, E. P., & Lidz, C. W. (1995). Conditional prediction: A model for research on dangerousness to others in a new era. International Journal of Law and Psychiatry, 18, 129-143.

National lnstitute of Justice. (1997). A study of homicide in eight US cities: An NIJ intramural research project. Washington, DC: US Department of Justice.

O'Sullivan, C. (1998). Ladykillers: Similarities and divergence's between gang-rapists and wifebatterers. In L, H. Bowker (Ed,), Masculinities and Violence, pp. 82-110. Thousand Oaks, CA:

Sage.

O'Sullivan, C., Wise, J., & Douglass, V. (1995, May). Domestic violence shelter residents in New York City: Profile needs, and alternatives to shelter. Report to the Mayor's Domestic Violence Commission.

Pataki,G. (1997). Intimate partner homicides in New York state Albany, NY: NY.

Pinard, G. F., & Pagani, L. (Eds.), (2000). Clinical Assessment of Dangerousness: Empirical Contributions. New York: Cambridge University Press.

Porter, C, & Villarruel, A. (1993). Nursing research with African American and Hispanic people: guidelines for action. Nursing Outlook 41 (2):59-67.

Quinsey, V., Harris, G. T., Rice, M. E., & Cormier, C. (1998). VioIent offenders: Appraising and managing risk (1st ed.). Washington, DC: American Psychological Association.

Roehl, J., & Guertin, K. (1998). Current use of dangerousness assessments in sentencing domestic violence offenders. Pacific Grove, CA: State Justice Institute.

Roehl, J. & Guertin, K. (2000). Intimate partner violence: The current use of risk assessments in sentencing offenders: Justice System Journal, *21* (2), 171-198.

Rosenfeld, R. (1997). Changing relationships between men and women. A note on the decline of intimate partner homicide. Homicide Studies, 1(1), 72-83.

Saunders, D. G. (1995). Prediction of wife assault. In J. C. Campbell (Ed.), Assessing the risk of dangerousness: Potential for further violence of sexual offenders batterers and child abusers (pp. 68-95). Newbury Park, CA: Sage.

Sharps, P. W., Campbell, J. C., Campbell, D. W., Gary, F. A., & Webster, D. W., (2003). Risky mix: drinking, drug use, and homicide. *NIJ Journal* (250):8-13.

Sharps, P. W., Campbell, J. C., Campbell, D. W., Gary, F. A., & Webster, D. W. The Role of Alcohol Use in Intimate Partner Femicide. *The American Journal on Addictions* 10:122-135, 2001a.

Sharps, P. W., Koziol-McLain, J., Campbell, J. C., McFarlane, J., Sachs, C. J., & Xu, X. (2001b). Health care provider's missed opportunities for preventing femicide. *Prev.Med.* 33:373-380.

Sheridan, D. J. (1998). Measuring harassment of abused women: A nursing concern. Oregon Health Sciences University School of Nursing: Doctoral dissertation.

Silva, C., McFarlane, J., Soeken, K., Parker, B., & Reel, S. (1997). Symptoms of post-traumatic stress disorder in abused women in a primary care setting. Journal of Women's Health, 6(5), 543-552.

Smith, M. N., & Brewer, V. E. (1990). Female status and the "Gender Gap" in US homicide victimization. Violence Against Women, 1(4), 339-350.

Smith, P. H., Moracco, K. E., & Butts, J. (1998). Partner homicide in context: a population based perspective. Homicide Studies, 2, 400-421.

Stack, S. (1997). Homicide followed by suicide: An analysis of Chicago data. Criminology 35(3),435-453.

Straus, M. A., & Gelles, R. J. (1990). Physical violence in American families: Risk factors and adaptations to family violence in 8.145 families. New Brunswick: Transaction Publishers.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1995). The Personal and Relationship Profile PRP: A Package of Instruments for Research and Clinical Screening of Couple Violence. Durham, NH: Family Research Laboratory, University of New Hampshire.

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

Sullivan, M.C., Rumptz, M.H., Campbell, R., Eby, K.K. & Davidson, W.S. (1996) Retaining participants In Longitudinal Community Research: A Comprehensive Protocol. Journal of Applied Behavioral Science32 (3):262-276.

Tanaka, J. S. (1987). How big is big enough? Sample size and goodness of fit in structural equation models with latent variables. Child Development, 58, 134-146.

Tjaden, P., & Thoennes, N. (1998, April). Stalking in America: Findings from the National Violence Against Women Survey. Research in Brief from National Institute of Justice/Centers for Disease Control and Prevention. Washington, D.C.: US Department of Justice.

Tjaden, P., & Thoennes, N. (1998, November). Prevalence. Incidence and Consequences of Violence Against Women: Findings from the National Violence Against Women Survey. Research in Brief from National Institute of Justice/Centers for Disease Control and Prevention. Washington, D.C.: US Department of Justice.

Tolman, R. M., & Bennett, L. (1990). A review of quantitative research on men who batter. Journal of Interpersonal Violence 5 (1),87-118.

Walker, L. E. (1984). The battered woman syndrome. New York: Springer.

Websdale, N. (1999). Understanding domestic homicide Boston: Northeastern University Press.

Webster, C. D., Harris, G. T., Rice, M. E., Cormier, C., & Quinsey, V. L. (1994). The violence prediction scheme. Assessing dangerousness in high risk men. Toronto, Canada: Centre of Criminology, University of Toronto.

Weiner, N. A., Zahn, M. A., & Sagi, R. J. (1990). Violence: Patterns, causes, public policy. Harcourt Brace Jovanich.

Weisz, A., Tolman, R. & Saunders, D. G. (2000). Assessing the risk of severe domestic violence. Journal of Interpersonal Violence 15 (1),75-90.

West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.H. Hoyle (Ed.), Structural equation modeling: Concepts, issues. and applications, (pp. 56-75). Thousand Oaks, CA: Sage.

Williams, K. (1999). Preliminary data using the DVI to predict reassault in Colorado. Paper presented at the New Frontiers in Risk Assessment Conference. Vancouver, BC, Canada.

Williams, K. R., & Houghton, A. B. (2004). Assessing the risk of domestic violence re-offending: A validation study. *Law and Human Behavior*: 437-455.

Wilson, M., & Daly, M. (1993). Spousal homicide risk and estrangement. Violence and

Victims. 8(1), 3-15.

Wilson, M., Johnson, H., & Daly, M. (1995). Lethal and nonlethal violence against wives. Canadian Journal of Criminology. 37, 331-362.

Wilt, S. A., Illman, S. M., & Brodyfield, M. (1995). Female homicide victims in New York City. New York: New York City Dept of Health.

Wolfgang, M. E. (1958). Patterns in criminal homicide. Philadelphia, PA: University of Pennsylvania Press.

Woods, S. J. (1999). Normative beliefs regarding the maintenance of intimate relationships among abused and non-abused women. *Journal of Interpersonal Violence* 14 (5), 479-491.

Zawitz, M. W. (1994). Violence between intimates. Washington, DC: Bureau of Justice Statistics.