The effects of a short-term batterer treatment program for detained arrestees: A randomized experiment in the Sacramento County, California Jail *

Bruce G. Taylor, Ph.D.
Director of Research
Police Foundation

and

Christopher D. Maxwell, Ph.D.
Associate Professor & Research Scientist
Michigan State University & the University of Michigan

Final Report submitted to

Department of Justice
Crime and Violence Prevention Center
1300 I Street, Suite 1150
Sacramento, CA 95814
&
National Institute of Justice
Office for Justice Programs
U. S. Department of Justice

* This study was supported by the National Institute of Justice (Grant # 98-IJ-CX-K014). The views expressed are those of the authors and do not necessarily represent the official positions of the California Department of Justice, the National Institute of Justice, the Police Foundation, Michigan State University, or the University of Michigan. Direct all correspondence to Dr. Bruce Taylor at taylorzzz@msn.com.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

BACKGROUND

There have been many evaluations of batterer education programs. There are at least eight published literature reviews of over 40 single-site evaluations. However, most of these evaluations have lacked sufficient methodological rigor to yield conclusive answers about the effectiveness of batterer education programs. In fact, there have been only a handful of investigations that can make any legitimate claims about differences between treated and untreated batterers. Even scarcer are programs for batterers arrested for a domestic violence (DV) offense awaiting adjudication. While there has been a proliferation of community-based programs for offenders convicted of DV offenses, throughout the country, few programs have been developed for DV offenders awaiting trial and little to no evaluation research has been conducted on these programs. The absence of these types of programs is worrisome, for many DV arrestees are able to pressure their victims into dropping the charges. In turn, these batterers do not get convicted and do not receive treatment.

METHODS

This study evaluates the effects of an early intervention program to provide DV education for DV arrestees on reducing repeat DV through a true experimental design. Based in the Sacramento County Sheriff's Department's Main Jail, the research randomly assigned over six hundred batterers to either the batterer treatment wing of the jail or to a no-treatment control group in another wing of the jail. The objectives of the research were to evaluate the effectiveness of this program against a no-treatment control group on reductions in DV recidivism.

The persons eligible to enter the experiment included: (1) adult males, (2) arrested for a misdemeanor or felony DV offense in Sacramento County, California against a female victim, (3) who did not put forward a $ 10,000 bail, (4) have no criminal gang affiliations, and (5) have no prior history of serving a sentence in jail or prison for one year or greater. Data were collected for about 10 months on 629 arrest cases for male perpetrated DV in Sacramento County, California from September 27, 1999 to August 6, 2000.

Interviews were conducted with the victims and batterers shortly after the arrest that placed the batterer in the Sacramento jail, and again six months post-arrest. Additionally, official police arrest data on recidivism were collected and analyzed for up to one year post-arrest. In an effort to standardize the results, this experiment utilized the same three measures of recidivism that were used in the six experiments that made up the Spouse Assault Replication Project (SARP) funded by the National Institute of Justice (NIJ) (Sherman, 1992b). The SARP measures of recidivism the study used included: Time to failure (i.e., the amount of time between the conclusion of the treatment/control condition and any new repeat offenses), prevalence of failure (i.e., the proportion of batterers that committed new offenses), and incidence/frequency of

failure. The study also included a fourth measure, severity of failure (related to the use of controlling behavior, psychological abuse, threats of physical assault, or actual physical and sexual assault). The study used a modified version of the Conflict Tactics Scale II (Straus et al., 1996) that has sub-scales for controlling behavior, psychological abuse, threats of physical assault, actual physical and sexual assault and injuries. Dichotomous prevalence measures of violence were examined through logistic regression modeling. Time-to-failure measures were analyzed with proportional hazard survival analysis. The incidence or number of violent acts measure was examined through a negative binomial regression/count model.

RESULTS

Arrestee/6-month follow-up measures of DV recidivism: For both the treatment and control group the highest proportion of DV, based on batterer self-reports, is for psychological abuse. For both the treatment and control groups over 43 percent of the men admitted to psychological abuse, over 4 percent of the men in the whole sample admitted to threats of physical abuse, over 10 percent of the men in the whole sample admitted to physical abuse, over 21 percent of the men in the whole sample admitted to controlling behavior and over 49 percent of the men in the whole sample admitted to at least one of these four types of acts of DV.

The first finding was for the batterer 6-month follow-up self-report data for controlling behavior. The treatment variables in the logistic, negative binomial and Cox regression models were all statistically significant and in the predicted direction. The six-month Prevalence Model shows that the treatment group had a lower proportion of participants using controlling behavior than the control group ($b=-0.99$, $p<.05$). The difference or effect size is also substantively large. The model estimated that 63 percent fewer than the expected number of subjects assigned to the treatment group committed a new incident of controlling behavior against the victim. The six-month Frequency/Rate Model shows that the treatment group had a lower incidence or frequency of controlling behavior than the control group ($b=-0.65$, $p<.001$). The Frequency Model estimated that the treatment group committed 48 percent fewer than the expected frequency of incidents of controlling behavior against the victim. The six-month Time-to-Failure Model shows that there was a significant decrease in the hazard rate for the treatment group compared with the control group ($b = -1.33$; $p < .01$). In other words, the men assigned to the treatment group had a 74 percent reduction in the likelihood of having a new incident of controlling behavior, thereby creating a longer safety period of non-controlling behavior for the victim.

The next finding was for the batterer 6-month follow-up self-report data for any acts of DV. The treatment variables in the logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure of batterer self-reported acts of any DV.

The final finding was for the batterer 6-month follow-up self-report data for physical abuse. The treatment variables in the logistic, negative binomial and Cox regression models were

all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure for batterer reported acts of physical abuse.

Arrestee/batterer based 6-month follow-up measures of alcohol and drug use: Two statistically significant differences were observed in batterer self-reported alcohol and marijuana use. Sixty percent of the treatment group consumed alcohol within the six-month period after the intervention compared to more than 77 percent for the control group. Seventeen percent of the treatment group of batterers reported using marijuana within the six-month period after the intervention compared to 30 percent for the control group. While none of the other drug measures (related to other drugs, frequency or dependency) were statistically significant, it does suggest that the intervention was creating some type of changes in the treatment group.

Victim-based accounts of DV recidivism results: For both the treatment and control groups about 45 percent of the sample of victims reported experiencing controlling behavior from the men in the sample, over 60 percent of the sample of victims experienced psychological abuse from the men in the sample, over 23 percent of the sample of victims experienced threats of physical abuse from the men in the sample, over 30 percent of the sample of victims experienced physical abuse from the men in the sample, over 45 percent of the sample of victims experienced controlling behavior from the men in the sample, and over 67 percent of the victims experienced at least one of these four types of DV from the men in the sample.

In contrast to our finding suggesting the treatment program is reducing controlling behavior, the treatment variable in the victim logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure for victim reported acts of controlling behavior by the perpetrator. Also, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure of victim self-reported acts of any DV. Finally, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure for victim self-reported acts of physical abuse.

Police DV arrest results: The treatment and the control groups both had over 65 percent of the men in the sample re-arrested for a new DV offense at six months post intervention. The treatment and the control groups both had on average less than 1 re-arrest for a new DV offense at six months post intervention. The treatment and the control groups both had over 68 percent of the men in the sample re-arrested for a new DV offense at twelve months post intervention. The treatment and the control groups both had on average 1 re-arrest for a new DV offense at six months post intervention. The treatment variable in the 6 and 12-month regression models were all non-significant. That is, there was no difference between the treatment and control group in 6 and 12-month prevalence, frequency, and time-to-failure of new arrests.

CONCLUSION

Overall, what emerges from the analysis is that the batterer treatment program had an effect on the least serious form of DV (controlling behavior prevalence, frequency and time-to-

failure), based on the batterer self-reports (although no difference was observed based on the victim-based measure of controlling behavior). This is not to suggest that controlling behavior is not a serious problem. The controlling behavior subscale includes questions on keeping the victim from talking on the phone, spending time with friends, stopping her from going some place she wanted to go, stalking behavior, keeping her from using the family income, stealing money from her, taking away or threatening to take away her children, harming or threatening to harm her children, seriously harming or threatening to seriously harm other people she knows or suicidal type of behavior. Few would argue that this type of behavior does not represent serious violations of other's rights.

Although the victim-based measure of controlling behavior found no differences between the treatment and control groups, the victim data on this issue might be less instructive then the arrestee data. The nature of some of the controlling behavior items might make it more likely for the batterers to have a more accurate assessment. For example, the batterer would be in the best position to know if they are stalking the victim. That is, the victim might not realize that the batterer has stopped stalking them or miss that the batterer has increased their stalking behavior. A similar case could be made for other items on the controlling behavior scale such as harming the victim's children or victim's friends, stealing money from the victim, or hiding family income. Another possibility is that this finding is just an anomaly. However, this seems very unlikely due to the fact that all three aspects of this measure are significant (prevalence, frequency and time-to-failure). Also, the batterer treatment participants admitted to other types of more serious behavior like psychological and physical abuse at the same rate as the control group participants. If the batterer treatment participants were lying why would they have been untruthful only about the controlling behavior and not the other more serious forms of violence? Although not the focus of this study's intervention, one of the components of the Sacramento batterer treatment program was alcohol and drug education/prevention through a daily Twelve-Step Drug/Alcohol addiction support group. The treatment group had lower rates of marijuana and alcohol use further suggesting that something has occurred with the treatment group

To date, this study has provided one of the most rigorous tests of the effectiveness of jail-based batterer treatment. The experiment was implemented well. There were no overrides in this study (i.e., where the assigned condition to treatment or control group is bypassed in favor of something else). This study addressed some methodological problems found in prior research, such as disentangling the effects of treatment from sample selection effects, including a large sample of treatment and control group cases, studying the total population to be evaluated (as opposed to only those who completed the program), and proper specification of the primary outcome measures of DV using reliable measures. Also, a number of tests were conducted that suggest that potential biases that might have been introduced into this study did not have a measurable effect on the substantive results of this study. At a minimum the Sacramento program is not making things worse for the victims and some of the evidence suggests that it might in fact be doing some good in reducing controlling behavior and alcohol and drug use.

1. INTRODUCTION

Over the past few decades there have been many evaluations of batterer education programs. There are at least eight published literature reviews of over 40 single-site evaluations. However, most of these evaluations have lacked sufficient methodological rigor to yield conclusive answers about the effectiveness of batterer education programs. In fact, there have been only a handful of investigations that can make any legitimate claims about differences between treated and untreated batterers. Over the past five years rigorous experimental evaluations have been completed on traditional court-mandated batterer education programs.

Even scarcer are programs for batterers arrested for a domestic violence (DV) offense awaiting adjudication. While there has been a proliferation of community-based programs for offenders convicted of DV offenses, throughout the country, few programs have been developed for DV offenders awaiting trial and little to no evaluation research has been conducted on these programs. The absence of these types of programs is worrisome, for many DV arrestees are able to pressure their victims into dropping the charges. In turn, these batterers do not get convicted and do not receive treatment.

In 1995 a unique program for detained arrestees was developed in the Sacramento Sheriff's Department's Main Jail. The program was setup as an early intervention program to provide DV education for DV arrestees during their time of detention before going to court.

This study evaluates the effects of the Sacramento program on reducing repeat violence through a true experimental design. The research randomly assigned over six hundred batterers to either the batterer treatment wing of the jail or to a no-treatment control group in another wing of the jail. The objectives of the research were to evaluate the effectiveness of this program against a no-treatment control group on reductions in domestic violence recidivism. Interviews were conducted with the victims and batterers shortly after the arrest that placed the batterer in the Sacramento jail, and again six months post-arrest. Additionally, official police arrest data on recidivism are analyzed for up to one year post-arrest.

2. BATTERER TREATMENT RESEARCH LITERATURE

Therapeutic treatment programs for batterers became a popular court sanction beginning in the early 1980s. As state and national policies increasingly promoted arrest and prosecution to control domestic violence (Buzawa and Buzawa 1996), this sanctioning method has taken on greater importance. These changes meant that criminal courts needed to sanction an expanding pool of batterers, and judges began to rely upon group treatment programs as their sanction of choice despite victims' desires or willingness to cooperate (Rebovich 1996; Hanna 1996). By the

late 1990s nearly every state used batterer treatment programs and administrators estimated that nearly 80 percent of their clients were court-referred (Healey et al., 1998).

2.1 Extent of problem

Because of the substantial growth in the number of batterer treatment programs, it is important to understand how effective these programs are in changing batterers' behavior. The wide prevalence of the problem of DV has been well documented (see Tjaden and Thoennes, 1998). Domestic violence is defined as assaultive behavior involving adults who are married, cohabitating, or who have an ongoing or prior intimate relationship (Goolkasian, 1986). Research indicates just how pervasive this problem is today. Based upon crimes reported to the police in 1998, intimate partner homicides accounted for about 11% of all murders in the U.S. (Rennison and Welchans, 2000). Over four million women were victims of a violent crime in the year 2000 alone, and most of them knew their assailants, with a large proportion of these offenders being intimate partners of these women (BJS, 2001). In one of the most rigorous studies on this issue, Tjaden and Thoennes (1998) found that physical assault is widespread among American women and estimated that approximately 1.9 million women are physically assaulted annually in the United States. Research also suggests that once victimized a domestic violence victim is at even greater risk of experiencing future violence (Hilberman 1980; Hirschel and Hutchinson, 1992; Langan and Innes, 1986). In addition, violent homes are a risk factor for producing violent adults, thereby continuing the "cycle of violence" (Brisson, 1981; Dutton, 1988; Widom, 1992). Also, domestic violence has a heavy economic component, estimated at approximately $67 billion per year (Miller, Cohen, and Wiersema, 1996). Therefore, an intervention that reduces the likelihood of domestic violence will benefit many women and the nation as a whole. Also, since many victims stay with their partners even after the batterer's arrest and conviction, it is essential to use effective programs that can change abusive behaviors rather than simply delay it during a period of incarceration. And some have argued that batterer treatment groups have the potential to create a "ripple effect" throughout the criminal justice system (Dutton 1986).

These figures have major implications for the criminal justice system and have shaped the legal response to DV. Pro-arrest police policies have been promoted by advocates and widely adopted by police departments across the country (Buzawa and Buzawa, 1996). Increasingly, prosecutors as well have removed discretion traditionally given victims of DV and insisted that cases be pursued to conviction regardless of victim desires or willingness to cooperate (Rebovich, 1996; Hanna, 1996). These changes have meant that criminal courts have had to sanction an expanding pool of batterers, and they have increasingly come to rely upon batterer intervention programs as the sanction of choice (Hanson, 2002; Healey and Smith, 1998), with as many as one-third of convicted domestic violence offenders ordered by the courts to attend batterer treatment groups (Healey and Smith, 1998).

2.2 What are batterer intervention programs?

One of the earliest responses to domestic violence was the use of battered women shelters (Johnson and Kanzler, 1993). However, shelters alone could not deal with the phenomena of domestic violence victims returning and reestablishing relationships with the batterers (Hamberger and Hastings, 1993; Jennings, 1987; Snyder and Scheer, 1981), or the batterers picking new partners to abuse (Farley and Magill, 1988; Gondolf, 1987). Consequently, the field started to direct its attention to changing the behavior of the batterer (Feazell, Mayers, and Deschner, 1984).

Batterer treatment groups for men have been available in the United States since the late 1970s. The first batterer groups emerged as a reaction to prompting from victim advocates involved in the shelter movement and local men's groups. Although batterers were being seen in individual counseling and marital therapy, victim advocates believed that a different type of program was needed. Group treatment is believed by many to be more appropriate than individual treatment because it expands the social networks of batterers to include other men who are supportive of being non-abusive (Crowell and Burgess, 1996). The earliest batterer groups were mostly consciousness-raising groups to promote an anti-sexist message (e.g., AMMEND in Denver and EMERGE in Boston), but they gradually incorporated cognitive and behavioral therapy techniques (Gondolf, 1995). By the mid-1980s, skill-building and brief therapies developed by clinicians were also adapted by batterer groups (Gondolf, 1995). These early programs were fairly unstructured groups, working with batterers through a combination of consciousness-raising and peer self-help (Adams and McCormick, 1982; Johnson and Kanzler, 1993). Most of these early programs also were implemented within a context of feminist theory that spoke of men's need to control women.

Over the years, more structured small groups using psychoeducational and/or cognitive behavioral techniques aimed at ending violent behavior have replaced these early groups (Pirog-Good and Stets-Kealey, 1985), and generally still within a feminist context (Healey, Smith, and O'Sullivan, 1998;Jennings, 1987). While program curricula are fairly universal, the format and structure of batterer intervention programs (BIPs) often differ across programs. Program length and intensity can vary from a few days to one year, with one or two sessions per week. Other variations include groups being structured or unstructured, psycho-educational or self-help. Some of the goals of BIPs include justice and accountability (80% of BIP participants are referred by criminal courts), victim safety, and rehabilitation and behavioral changes such as skill building, attitude change, and emotional development (Bennett and Williams, 2001).

There are more than three-dozen evaluations of batterer treatment programs (see Babcock, Green and Robie, in press; Davis and Taylor, 1999, for reviews). Babcock, Green and Robie (In press) conducted a meta-analytic review of the findings of 22 studies evaluating treatment efficacy for domestically violent males. Babcock et al. (In press) reviewed the outcome literature of controlled quasi-experimental and experimental studies on the relative effectiveness of the Duluth Model, cognitive behavioral therapy (CBT), and other types of treatment on subsequent recidivism of violence. Study design and type of treatment were tested as moderators. Treatment design tended to have a small influence on effect size. There were no differences in effect sizes in comparing Duluth Model versus CBT-type interventions. Overall, the effects due to treatment were in the small range, with the interventions having a minimal impact on reducing recidivism beyond the effect of being arrested.

While evaluations of BIPs have shown an evolution toward more rigorous science since the first studies in the early 1980s, many still lack sufficient methodological rigor to yield valid answers about the effectiveness of the programs. This study in Sacramento is one of the most recent attempts to test batterer treatment using an experimental design that randomly assigns men arrested for a domestic violence offense to batterer treatment or to a no-treatment control condition. This study addresses some methodological problems found in prior research (see comments by Hamberger and Hastings, 1993; Palmer, Brown, and Barrera, 1992), such as disentangling the effects of treatment from sample selection effects, including a large sample of treatment and control group cases, studying the total population to be evaluated (as opposed to only those who completed the program), and proper specification of the primary outcome measures of domestic violence using reliable measures.

At least nine published reviews of more than forty published single-site evaluations (e.g., Babcock, Green and Robie, in press; Eisikovits and Edleson, 1989; Gondolf, 1995; Rosenfield, 1992; Saunders, 1996; Tolman and Bennett, 1990) and nine research reviews (e.g., Davis and Taylor, 1999; Hamberger and Hastings, 1993; Crowell and Burgess, 1996; Dobash, et al., 1995; Dutton, 1988; Rosenbaum and O'Leary, 1986; Saunders and Azar, 1989; Tolman and Edelson, 1995) show there is no lack of empirical studies on batterer treatment programs. This is deceptive, however, since only a handful of investigations can make any valid claim about differences between treated batterers and untreated batterers. In the following section there is a review of the three generations of batterer treatment studies: (1) Studies that failed to use control groups and only examined batterers assigned to treatment programs, (2) quasi-experiments, and (3) studies which randomly assigned batterers to treatment.

2.3 Studies without control groups

The oldest and largest portion of empirical literature is studies that only include batterers assigned to treatment programs. Included in this set of studies are: (i) studies which assess violence or other outcomes among participants in treatment programs only after treatment (single

group post-test-only designs), (ii) studies that measure violence among treatment participants both before and after treatment (single group pretest-post-test designs), and (iii) studies which compare violence of batterers who complete treatment with batterers who did not complete treatment, but were assigned it. This literature contains more than two-dozen examples. Gondolf (1999a, 1999b) has conducted one of the best examples of this design. Gondolf (1999a, 1999b) followed 840 men in four different BIPs in Pittsburgh, Houston, Dallas, and Denver, using both court records and partner interviews, for as long as three years after intake. Despite differences in referral (diversion v. conviction), length of treatment (3 to 9 months), or additional services (e.g. mental health, substance abuse) there were no significant differences between the four groups at 15-month follow-up in re-assault (32%), controlling behavior (45%), re-arrest (26%), or victim perception of safety (72%). These studies were important in developing this area of research, but they are difficult to interpret.

Studies without any treatment control groups provide no reference point to judge whether treatment programs reduce violence. Single group pre-test/post-test designs that show a reduction in violence after batterers participate in a treatment program are problematic because research repeatedly has shown that domestic violence declines after the police are called, even if nothing else is done. In fact, research suggests that only about a third of batterers commit another act of domestic violence within six months after the police intervene (e.g., Davis and Taylor, 1997; Fagan, et al., 1984; Maxwell, 1998). In addition, among studies that compared outcomes between batterers who complete treatment and those who do not, the treated and untreated (dropout) groups are not comparable prior to treatment. Palmer, Brown, and Barrera (1992) suggest better attendance indicates a greater motivation to change, even before treatment.

2.4 Quasi-experiments

The next evaluation studies include quasi-experimental designs using non-equivalent-matched groups. In at least four studies, batterers mandated to treatment were compared with batterers who received other interventions (Harrell, 1991; Chen, et al., 1989; Dutton, 1986; Dobash, et al., 1996). These studies have also generally been more rigorous than those without control groups (e.g., by including larger sample sizes, not relying only on batterer self-reports to determine new violence, and having at least one-year follow-up periods).

In the Chen, Bersani, and Denton (1989) study they compared 120 convicted batterers in an 8-session BIP with 101 convicted batterers who did not get the BIP. Participants in BIP were half as likely to re-offend during the year after treatment, but only if they attended at least 6 of the 8 sessions.

Dutton (1986) compared police information records of 50 Vancouver-area men who received a 16-week cognitive behavioral BIP with the records of 50 men who were excluded from the BIP for some reason. Men in the BIP and men in the comparison group did not differ either demographically or on pre-conviction records of violence. Results suggest a statistically significant effect of BIPs compared to the alternative condition.

Dobash, Emerson-Dobash, Cavanagh and Lewis (1996) reported on a quasi-experiment evaluating a treatment program in Great Britain. Dobash et al. (1996) examined 256 domestic violence cases from sheriffs' courts in Scotland in which defendants were sentenced to batterer treatment or to another sentence (probation, court supervision, or prison). Few details are given about how the control group was selected, but the authors note that batterers in the treatment group were significantly older and more likely to be employed than batterers in the control group. According to court reports at 12 months follow-up, 7% of the treatment group recidivated compared to 10% of the control group. Data from victim surveys indicated that half as many batterers assigned to treatment committed new violence at three or 12 months as controls. These two comparisons are reported to be statistically significant. However, the success rate for interviews was low: Dobash et al. (1996) interviewed only 43% of the victims at the first follow-up interview, 34% at the second interview, and 25% at the third interview.

Among the four quasi-experimental studies, three reported lower violence among treated batterers than among untreated batterers (Chen et al. 1989; Dutton 1986; Dobash, et al. 1996). However, the effect sizes were not always statistically significant, and they depended upon which outcome measures were examined and whether comparisons involved all men assigned to treatment or only those who completed a requisite number of sessions. In contrast to these three studies, Harrell (1991) found that men who completed treatment were more abusive afterward compared to a control sample. Harrell (1991) studied batterers ordered to one of three 12-session BIPs, and batterers not ordered to a BIP. The BIPs all differed in their orientation to services. At follow-up, men in the BIP condition (n=81) were more likely to have been physically violent and more likely to have been re-arrested than men in the control condition (n=112).

Although quasi-experiments are a step forward from studies without control groups, as pointed out by Palmer et al. (1992), they do not provide unbiased estimates of treatment effects because we cannot know whether batterers assigned to treatment and those in control groups are equivalent prior to the treatment. In three of the quasi-experiments (Dutton [1986], Harrell [1991], and Dobash, et al. [1996]), the control group comprised batterers more prone to recidivate than those in the treated group and so the study favored finding treatment effects. Researchers may try to statistically control for any pretreatment differences between groups; however, these controls may not include key unobserved or unmeasured factors related to outcomes.

2.5 Randomized experiments

The safest way to ensure that sample mean estimates are unbiased is by randomly assigning batterers to treatment and control groups. In 1992 Palmer et al. conducted the first batterer's treatment experiment, randomly assigning batterers to a true no-treatment control group. Fifty-nine probationers in Canada were assigned using a block random procedure to either a ten-session psycho-educational group (combining group discussion with information) or a no-treatment control group. To measure outcomes, Palmer and her colleagues (1992) examined police reports six months after treatment and found recidivism rates (domestic physical abuse or serious threats) for the treatment group to be just one-third that of the control group. Even with the small N, this difference was statistically significant. While Palmer et al. (1992) attempted to measure violence based on surveys of victims and batterers, low response rates and a small sample size precluded any analysis of recidivism based upon interview data.

Edleson and Syers (1990) randomly assigned 283 batterers to one of three programs in Minneapolis (self-help versus educational v. combined) and one of two program intensities (weekly for three months versus twice weekly for four months). However, there was no true (no treatment) control group included in this study. A six-month follow-up with 92 program completers and their partners found no significant differences between models or intensities, although there was a non-significant trend favoring the educational approaches over the self-help approach.

Saunders (1996) randomly assigned 218 batterers in Madison, Wisconsin to cognitive-behavioral or process-psychodynamic group interventions. This study also did not include a control group. In 18 to 54-month follow-up with program completers, there were no differences in arrests or in victim-reported violence or fear of violence between the two treatment approaches. However, men who had higher levels of dependency did better in the psychodynamic treatment, and men who had a more antisocial orientation did better in the cognitive-behavioral program.

Dunford (2000) compared outcomes for U.S. Navy men randomly assigned to (a) a 1-year cognitive-behavioral BIP, (b) a 1-year couples group, (c) a rigorous monitoring program similar to assertive probation work, or (d) a safety planning condition approximating a control group. Men with substance abuse problems or mental disorders were excluded from the study. At one-year follow-up, 48 men in the BIP condition had slightly lower incidence of recidivism by partner report than 50 men in the control group (but the difference was not statistically significant), and there were no differences in one-year re-arrest rates across the groups.

Feder and Forde (2000) studied 404 men in Broward County, Florida randomly assigned to probation plus a Duluth-based BIP or probation only. At follow-up, there were no significant differences between the BIP and the probation-only group in self-reported domestic violence offending. At one-year follow-up, men in the BIPs were no less likely than men in the control group to be re-arrested for domestic violence.

Taylor, Davis, and Maxwell (2001) compared men in Brooklyn, New York randomly assigned to either a 40-hour Duluth-based BIP (n=186) or 40 hours of community service (n=190). Follow-up data collection occurred at 6 and 12 months, and included both official records of complaints/arrests and victim interviews. Results showed significantly lower recidivism for men in BIPs compared to the control group using official arrest records, but no difference between BIPs and the control group using victim reports.

As seen with other studies, the experimental design studies show mixed results with a couple showing positive effects of treatment and others showing no effects. However, these studies were conducted with different populations that may account for part of the conflicting findings. For example, one study was done with men in the U.S. Navy, others were done with a broad group of DV offenders in the criminal justice system, and others with a narrower group of more motivated offenders from the criminal justice system.

3. DESCRIPTION OF THE TREATMENT USED IN THIS STUDY AND THE FIDELITY OF THE INTERVENTION

A system of careful monitoring of the treatment and control conditions for this experiment were established through a joint effort with the Sacramento County Jail staff. Each of the below described elements of the intervention were implemented completely. The Sacramento Jail batterer intervention program became operational in March 1997. In Sacramento, at the time of this study, it took DV arrestees several months to go from arrest to sentencing. California law, at the time of this study, did not allow alleged batterers to be released on their own recognizance, arrestees had to post a $10,000 bond. The consequence of this policy was that many arrestees charged with domestic violence offenses were detained in jail for at least one week, even cases where there was ultimately a dismissal. This program is as an early intervention program to provide treatment/education during this time of detention. The program was run from a special wing in the Sacramento jail, separate from the regular population of inmates. The program targeted defendants who were charged with DV offenses who had little or no prior criminal records (60% of the defendants in the Sacramento jail who were charged with DV offenses did not have prior criminal records). The program objectives were to raise awareness (“ break down the wall of denial” for DV offenders), educate program participants on

non-violent conflict resolution skills, increase receptivity to long-term treatment, provide the opportunity for drug and alcohol interventions, and reduce DV recidivism.

The activities of the program included: Mandatory detention in a special DV jail-wing supervised by correction officers who have received special DV training, batterer educational workshops, daily Twelve-Step Drug/Alcohol addiction support groups, and strict regulations on television watching (special non-violent education programs were the only available programs). Batterer education classes were held daily and the research team checked attendance logs. These classes were conducted by MANALIVE (Men Allied Nationally Against Living in Violent Environments) and by WEAVE (Women Escaping A Violent Environment) on a rotating basis. The classes were about three hours in length. The average length of incarceration for this group of DV arrestees was about 5 days that was spent in this special wing of the jail receiving the interventions. The selection of DV arrestees into this special wing of the jail was not a voluntary process. Also, the arrestees in this special wing were required to at least attend the program classes and NA/AA groups and sit quietly.

The program curriculum was based on the Duluth Model. This model, rooted in a feminist perspective, assumes that domestic violence is a by-product of male and female sex roles that result in an imbalance of power. The tested curriculum included: Defining domestic violence, understanding the historical and cultural aspects of domestic abuse, and reviewing criminal/legal issues. Through a combination of instruction and discussion/interaction, participants were encouraged to take responsibility for their anger, actions, and reactions.

The research team conducted announced and unannounced assessments of the fidelity of the interventions. A sampling of intervention sessions were observed, coded and compared to the written curriculum by two independent raters present in each of the sessions (inter-rater reliability exceeded 95%), with over a 90% agreement between what was occurring in the sessions compared to the curriculum. Focus groups with men in a sampling of five sessions at the beginning and end of the study revealed that the participants understood the intervention, reported changes in knowledge and attitudes, were paying attention, participating and giving the sessions a chance to have the desired effect.

While the average number of sessions attended for the 317 treatment cases was five days, a small number (n=91) of men received less than a full day of treatment sessions. In these 91 cases the men were able to make bail within the first day of detention (but not at the time of booking) and were released. Given that these 91 men received less of the intervention than the 226 other treatment recipients there is a concern that this might dilute the effects of the

intervention. Only two differences were observed between the 91 men and 226 men (see Table 2). The 226 men were statistically older (b= 0.04, p<.05) and reported statistically more physical abuse in their past history (b= 0.29, p<.01) than the 91 men released within the first day of their detention. To address this problem both age and past physical abuse history were added as covariates in all the substantive models. Also, it is worth noting that the control group also had a similar number of men who were released within their first day of detention (n=87 of 312 control cases). The similar proportion of early releases in the treatment and control groups creates a balance and effectively negates the effects of early release. Additionally, all of the substantive models were estimated with and without these 91 cases and no differences in the models were observed.

Control group participants were assigned to the regular part of the jail and received the usual incarceration experience of persons detained in the Sacramento County jail (including no treatment services). The key issue with the control condition was to avoid any possible contamination with the treatment group. Based on interviews with jail staff and the research teams own observations these is no evidence of any contamination – with the control group participants receiving no treatment–related interventions during their detention period.

Another issue the research team had to contend with was the possibility of men in the experiment getting re-arrested and re-entering the experiment (this occurred in 95 cases). Each time a person was entered into the experiment the research staff checked the experiment log for a previous entry. For the 95 repeat cases, the study team assigned them back to their original condition (47 treatment and 48 control cases) to maintain a clear distinction between the treatment and control groups. Many other men in our experiment were arrested during the study period, but just not in Sacramento County, not for a domestic violence offense or they were able to make bail on the subsequent domestic violence arrest. When the author compared our substantive models with and without these 95 repeat cases none of the results materially changed, and there was no discernable pattern to these 95 cases in terms of differences between the treatment and control groups.

4. METHODS

4.1 Experimental design

The research design for this evaluation was a randomized experimental design, one of the most rigorous designs available for assessing the effectiveness of interventions. A basic flaw in many of the prior batterer treatment studies has been the absence of comparable/equivalent control groups. Some of these studies did not use any control group at all (post-test only designs), some used one treatment group that acted as its own control group (through a pre-post-test design), and others have compared batterers who complete treatment with those who drop out (see Davis and Taylor, 1999 for a detailed review of these studies). Because these groups are self-selected, such studies cannot be expected to produce unbiased estimates of treatment effects. The best of the batterer treatment studies have attempted to draw comparison groups in ways that maximize the likelihood that they will be similar to the treatment group. For example, there have been four randomized experiments (Palmer, Brown, and Barrera, 1992; Davis and Taylor, 1997; Feder and Forde, 2000; Dunford, 2000), and four quasi-experiments with matched control groups (Chen, Bersani, Myers and Denton, 1989; Dobash et al., 1996; Dutton, 1986; Harrell, 1991). The problem with the quasi-experimental studies is that although differences that can be measured can be statistically controlled the many unmeasured variables related to the outcome variable (e.g., motivation to change) cannot be controlled. Also, research suggests that even the most sophisticated of quasi-experimental designs are unable to produce the same results as experimental designs. When experimental designs results are contrasted with results from other major designs and statistical alternatives, such as propensity analysis, different effect sizes are often found (e.g., Lalonde, 1986; and Fraker and Maynard, 1987). Even when no differences are observed between experiments and quasi-experiments on the same topic, the variation in results across the quasi-experiments is greater than across the experiments (Lipsey and Wilson, 1993).

With experimental designs the units of analysis are randomly assigned to treatment and control conditions. The randomization process refers to the use of a chance mechanism as the method for allocating treatments to experimental units (Gelber and Zelen, 1986). In this study, assignment to the batterers' intervention or control group was based on a random numbers table. Randomization of units produces similarity between the characteristics of subjects allotted to different treatments, within limits predictable by probability theory (Armitage, 1996). The principal advantages of randomization include: the elimination of conscious bias on part of the selection of units for the study; the elimination of unconscious bias caused by known factors affecting treatment or the outcomes of interest; making treatment groups similar on prognostic factors that might affect the outcome variables and attrition; and the straightforward nature of the analysis of experimental data (Gelber and Zelen, 1986).

Randomized field experiments are typically thought of as the best method for eliminating threats to internal validity in evaluating social policies and programs (Berk et al. 1985; Boruch, McSweeny, and Soderstrom 1978; Campbell 1969; Campbell and Stanley 1963; Dennis and Boruch 1989; Fairweather and Tornatzky 1977; Riecken et al. 1974). RCTs provide the best counterfactual describing what would have happened to treatment group participants if they had not been exposed to the treatment (Rubin, 1974; Holland, 1986). When implemented well, randomization ensures that, in the aggregate, the individuals assigned to one condition are equivalent in every possible way to those assigned to another condition, within the limits of statistical fluctuation. Thus, the only average difference between the two groups is that one group receives the treatment and the other does not. While random assignment cannot prevent all unintended factors or all unknown factors from affecting the outcome, it offers the highest degree of protection from bias by averaging the effects of these factors over all levels of the experimental treatment factors. When comparisons are made among levels of these factors, the bias caused by the unintended effects will tend to cancel and the true factor effect will remain (Mason, Gunst and Hess, 1989).

4.2 Experimental misassignments

A major problem in conducting randomized experiments in field settings is misassignments (Berk, et. al., 1988). That is, cases where the subjects do not receive the treatment or control condition that they are assigned to by the random assignment process. In the Davis and Taylor (1997) study misassignments/overrides were present in 14% of their cases. For 14% of the participants assigned to the control condition they were overridden to the batterer group. A number of procedures were put in place for this study to deal with this problem and there were no misassignments in this study. The key reason for the absence of overrides was the cooperation the research team received from the jail detention staff — from the highest levels to the line officers. The principal investigator met with senior Sheriff management about the experiment on several occasions before, during and after the study to gain their full support. Additionally, the principal investigator met with the line officers and supervisors for a full day of orientation to the study. Details of the study were reviewed and the officers were given a number of opportunities to provide input into the mechanics of intake for the study. By spending the extra time needed to get “buy in,” the integrity of the assignment process was held completely intact.

4.3 Human subject issues

Research staff interviewed victim and batterer participants at a baseline interview (held immediately after the arrest for the triggering event) and six-months after the intervention or control condition was concluded. The interviews were conducted over the phone, except for the

baseline batterer interviews that were done in the jail and for those who were not available for interviewing over the phone. The survey instruments all took on average about 40 to 45 minutes to administer by a trained interviewer. Strict human subjects standards for informed consent were followed for this research project. In order to ensure full voluntary subject participation, an extensive interviewer-training program was put in place. Interviewer training was a detailed process to ensure that the purposes and sponsorship of the research were explained thoroughly, that questions were asked in a non-threatening manner, that the procedures for confidentiality were clearly understood, and that interviewers were skilled in communicating these facts to potential respondents. Respondents were told the following: (1) their participation was optional and voluntary, (2) their responses were to be known only by the research team; (3) the interview responses and log books would be destroyed one year after the completion of the research project; and (4) each research participant was informed in detail about the nature of the study (its sponsors, sources of funding, objectives and goals, probable duration, and the expected length of time it would take to fully participate in the study). After the above issues were discussed with participants, their consent to participate in this study was solicited. Written or verbal informed consent was sought.

4.4 Measures

Recidivism/Violence (Victims, Batterers, and Official Records at all time waves): In an effort to standardize the results, this experiment utilized the same three measures of recidivism that were used in the six experiments that made up the Spouse Assault Replication Project (SARP) funded by the National Institute of Justice (NIJ) (Sherman, 1992b). The SARP measures of recidivism the study used included: Time to failure (i.e., the amount of time between the conclusion of the treatment/control condition and any new repeat offenses), prevalence of failure (i.e., the proportion of batterers that committed new offenses), and incidence/frequency of failure. The study also included a fourth measure, severity of failure (related to the use of controlling behavior, psychological abuse, threats of physical assault, or actual physical and sexual assault). Three sources of data were used: Victim, batterer, and police arrest reports. To assess self-reported violence the study used a modified version of the Conflict Tactics Scale II (Straus et al., 1996) that has sub-scales for controlling behavior, psychological abuse, threats of physical assault, actual physical and sexual assault and injuries.

The controlling behavior subscale includes questions on keeping the victim from talking on the phone, spending time with friends, stopping her from going some place she wanted to go, following her against her will, keeping her from using the family income, stealing money from her, taking away/threatening to take away her children, harming or threatening to harm her

children, seriously harming or threatening to seriously harm other people she knows or threatening to kill or hurt himself or actually hurting himself.

The psychological abuse subscale includes questions on swearing/screaming at the victim, accusing her of being involved with another man, putting her down/insulting/calling her names, throwing/smashing/hitting/kicking something, destroying any property, hurting any pets, and stomping out of the room during a disagreement with the victim.

The threats of physical assault subscale includes questions on threatening to hit, attack or harm the victim, threatening her with a knife or a gun, threaten to make her have oral/anal/vaginal sex and threatening to kill her.

The physical/sexual abuse subscale includes questions on pushing/grabbing/shoving the victim in anger, restraining her against her will, twisting her arm or pulling her hair, slamming her against a wall or other hard surface, slapping her, kicking her, punching her or hitting her with something that could hurt, throwing something with the intent to hurt her, stabbing her with a knife or other sharp object or shooting her with a gun, choking or strangling her, burning or scalding her on purpose, beating her up, attempting or committing any unwanted sexual contact such as grabbing or fondling her against her will, using force or threats of physical force to have oral/anal/vaginal sex with her against her will, and attempting to kill her.

Table 3 (in the Appendix) presents reliability statistics for each of the four scales of controlling behavior, psychological abuse, threats with physical abuse, and physical abuse. The victim survey measures for these four scales had Chronbach alpha scores ranging from .70 to .91. The batterer survey measures for these four scales had Chronbach alpha scores ranging from .64 to .80 (except for the frequency of threats of physical abuse measure which had a Chronbach's alpha score of .20). With the exception of the batter reported frequency of threats of physical abuse measure, all of the other measures had acceptable to high levels of reliability. In the case of the batter reported frequency of threats of physical abuse measure, the results will have to be viewed cautiously. The study staff also checked police records for new crime reports and arrests within 12 months of the arrest. While many domestic violence incidents are not reported to authorities, data from official records have been generally acknowledged as a useful supplement to victim reports of violence (e.g., Sherman, 1992). Because police data are available for most of the sample (only 7% or 47 of 629 cases were missing official record follow-up), selection bias is not an issue (as it is for victim and batterer self-reports).

Drug and Alcohol use (Batterers at baseline and six month follow-up): The study used the Arrestee Drug Abuse Monitoring (ADAM) self-reported drug use measures. The study measures included questions on the prevalence and frequency of substance use for alcohol, marijuana, crack cocaine, powder cocaine, heroin/dilaudid/morphine (opiates),

amphetamines/speed/methamphetamine, and other illegal drugs. Participants were first asked whether they used any of these substances (before and after their arrest), when was the first time they used since the arrest, how many days they used in the past 30 days, whether they have consciously tried to cut down or quit using on their own and whether they were successful, whether they went for inpatient or outpatient drug treatment, and whether they felt that they needed or were dependent on alcohol or a particular drug.

Demographics: A number of demographics measures were collected from official records for the entire sample and from self-reports from the victims and batterers (e.g., ethnicity, marital status, income, employment, education, and age).

Intervention Assignment: A number of measures were collected on whether the participant was assigned to the treatment or control group, how long they were held in jail, and (for the treatment group only) the number and type of sessions they attended while in jail. The control group was also asked questions about services received in jail, including whether they were at any time misassigned to the treatment group.

5. DATA ANALYSIS

Given that treatment assignment was done at the individual-level and that there was no dependence among the cases, all of the project analyses were carried out at the individual-level. By not having to deal with any clustering/dependency issues, the statistical power of the study is enhanced by conducting individual-level analyses.[1]

Dichotomous prevalence measures of failure were examined through logistic regression modeling, a regression technique designed for dependent variables coded as zero or one. Time-to-failure measures were analyzed with proportional hazard survival analysis (i.e., Cox regression). Survival analyses/Cox regression was used to examine differences between the treatment and control groups in elapsed time to first failure. The Cox regression model enables the efficient modeling of data in a multivariate context when the dependent measure is time censored (e.g., no case is followed for infinity). The incidence or number of failures measure is examined through a negative binomial regression/count model, testing whether the control group

[1] For example, school studies that randomly assign classes to interventions need to account for the clustering of students within classes through Hierarchical Level Modeling which reduces the statistical power of the study).

committed more frequent acts of new violence than the treatment group. Negative binomial regression models were developed specifically for the kind of distribution of failures evident with these data (i.e., a large portion of the sample did not fail at all during the time observed, some failed once, fewer failed twice, and a handful failed more often). This kind of highly skewed distribution seriously violates the normality assumption of OLS regression and ANOVA (even with log or other data transformations) and requires a negative binomial regression model (see Sherman and Smith, 1992 and Land, McCall, and Nagin, 1996). The negative binomial regression models were estimated using Stata software (version 8.2), with SPSS (version 12.0) being used for all the other analyses.

Although not strictly necessary because this study is using experimental data, a set of covariates has been added to the model. There are several reasons for introducing covariates even in analyzing data from randomized experiments. The introduction of covariates to the model improves the precision of the treatment comparisons and corrects for any major imbalances in the distribution of these covariates across the treatment and control group that may have occurred due to chance, adjusts for the natural variations between suspects within the two comparison groups, and allows testing for additional non-experimental hypotheses (e.g., effects of prior criminal record on new domestic violence).

5.1 Missing data issues

Given the longitudinal nature of this study, the problem of missing data was examined very carefully. There were two types of missing data that were examined: Cases of questionnaires only partially completed by respondents and cases lost to attrition. For the first situation, the effects of employing various imputation-based procedures to fill in missing values (mean imputation, regression imputation, and nonignorable missing-data models) were compared. This step specifically involves constructing a regression model that computes a predicted value for all cases based on those cases with valid data, and then uses this predicted value to replace the remaining missing data.

For the latter situation of attrition, a series of group comparisons (to determine whether cases lost to attrition differ from those who are retained) on key demographic variables (available for the entire sample from official records and baseline data) were conducted. Next, the research team employed Heckman's (1979) two-step process. The first step was to specify a model through the use of a multiple regression of the selection process that was captured in a single predicted latent measure. This latent measure was then entered into all the substantive outcome models as an independent variable to more fully specify the model.

Reviewers of the batterer treatment literature (e.g. Edleson, 1996; Gondolf, 1997) have stressed the importance of obtaining high response rates with respondents. Our follow-up attrition rate (about 80% for batterers and 75% for victims) clearly falls short of this goal. To investigate the impact of the attrition, a number of analyses were conducted and no measurable differences between the participants and non-participants were found. No systematic patterns emerged connected to the analyses of any of the missing data. The best indications are that the missing values are random. The saved selection process latent variables also proved to be non-significant and were unrelated to any of the outcome measures.

The police data was available for most of the sample (only 7% or 47 of 629 cases were missing official record follow-up), and all bivariate tests comparing the 47 cases without police data to the 582 cases with police data proved to be non-significant across a range of demographic/background/history variables (batterer age, race/ethnicity, education, employment, income, relationship status with victim, prevalence of self-reported psychological or physical abuse before arrest). A logistic regression of the 47 missing cases versus 582 cases with data also proved to produce only non-significant independent variable. Additionally, when the saved selection process latent variables were used in the regression models of the official record outcome measures they each also proved to be non-significant.

The batterer six-month follow-up self-report data was available for only small portion of the sample (about 20% or 119 of 629 cases had batterer six month follow-up self-report data). This stands in contrast to the batterer baseline self-report data where 560 (90%) of the 629 batterers completed the interviews. The difference in participation rates is attributable to the availability of the batterers for face-to-face interviews while in jail at the time of the baseline period. Most of the 560 batterers not interviewed could not be contacted/found after their release from jail. All bivariate tests comparing the 119 completed six-month interview cases to the 560 non-completed cases proved to be non-significant across a range of demographic/background/ history variables (batterer age, race/ethnicity, education, employment, income, relationship status with victim, prevalence of self-reported psychological or physical abuse before arrest). A logistic regression of the 560 missing cases versus 119 cases with data (see Table 2 in the Appendix) also proved to produce only non-significant independent variables, and when the saved selection process latent variables were used in the regression models of the batterer 6-month outcome measures they each also proved to be non-significant.

The victim six-month follow-up self-report data was available for only small portion of the sample (about one-quarter or 143 of 629 cases had victim six month follow-up self-report data). This is consistent with the victim baseline self-report data where only 208 (33%) of the

629 cases had completed interviews. In both interview waves the missing victim interview cases were due to the transient nature of this population- most of the victims the study team were able to contact agreed to be interviewed. The problem was that most of the victims were not residing in their last known address based on police records. All bivariate tests comparing the 143 completed six-month interview cases to the 486 non-completed cases proved to be non-significant across a range of demographic/background/history variables (batterer age, race/ethnicity, education, employment, income, relationship status with victim, prevalence of self-reported psychological or physical abuse before arrest). A logistic regression of the 486 missing cases versus 143 cases with data (see Table 2 in the Appendix) also proved to produce only non-significant independent variables, except for education variable ($b= 0.19$, $p<.05$) and the intercept. That is, controlling for all the other variables in Table 2, those with a completed six-month victim interview had a statistically significant higher level of education than those without a completed six-month victim interview. However, when the saved selection process latent variables were used in a regression model for the victim six-month outcome measures they each proved to be non-significant. As a safeguard for any potential bias that might be introduced by the presence of higher educated victims in the sample, this variable is included in the set of covariates for all of the substantive model tests.

Given that there are concerns that interviewed victims and batterers may constitute a different kind sample, it is reassuring that there was no significant effect from the interview measure. Thus, the above set of non-significant model findings lends some support to the idea that the victims and batterers who completed an interview were not significantly different than those who did not.

5.2 Sample Characteristics

The persons eligible to enter the experiment included: (1) adult males, (2) arrested for a misdemeanor or felony DV offense in Sacramento County, California against a female victim, (3) who did not put forward a $ 10,000 bail, (4) have no criminal gang affiliations, and (5) have no prior history of serving a sentence in jail or prison for one year or greater.

Data were collected for about 10 months on 629 arrest cases for male perpetrated domestic violence in Sacramento County, California from September 27, 1999 to August 6, 2000. During that same period, the Sacramento County population (based on the 2000 U.S. Census) was over one million people, with the Sacramento County Main Jail booking approximately 50,000 male arrestees. Of those 50,000 male arrestees, about 5% or 2,500 of those were arrested for DV offenses (with about 7 DV arrestees per day). The batterer program excluded 50% of the available pool of DV arrestees because of prior convictions. Therefore, the program's eligibility pool was 1,250 cases for the year (50% of 2,500 DV arrestees) or 1,040

cases for the 10-month study intake period. From the 1,040 cases the study lost an additional 370 cases due to arrestees making bail and 41 cases for a variety of other reasons (e.g., arrestees deemed to dangerous or intoxicated to be interviewed for a research project).

All the batterers were males, with an average age of thirty-three years old (see Table 1 in the Appendix). Thirty-four percent of the batterers reported being White, 28 percent reported being African American, 17 percent reported being some "other" race, 13 percent reported being a "mixed" race, 4 percent reported being American Indian, and 3% reported being Asian/Pacific Islander. In a separate question, 23% of the batterers reported being of Hispanic origin. About one third of the men reported not having a high school degree, another 39 percent reported earning a high-school diploma or GED, and 27% reported completing some formal education beyond high school. Approximately three-quarters (74 percent) of the men reported employment (either part-time or full-time) in the 30-day period before their arrest. The mean household income was approximately $26,000 per year, while the mean personal income level was about $18,000 per year. Looking at the time period immediately prior to the arrest, more than a third of the men (37 percent) reported being married to the victim, 40 percent reported living together with the victim (but not being married), 12 percent reported that their victim was a past intimate partner, 10 percent reported being in an intimate relationship with the victim but not living together, and 1 percent were separated or divorced from the victim. The vast majority of the men (88 percent) in the study reported perpetrating psychological abuse against the victim before the arrest and more than half being physical abusive (54 percent) before the arrest. Based on the treatment program criteria (excluding men with extensive prior histories with the criminal justice system) this sample did not have extensive arrest histories. However, more than one-third of the men had at least one prior arrest for domestic violence (36 percent) and nearly one-quarter of the men (24%) had at least one prior arrest for something other than domestic violence. Only one difference was observed between the treatment and control groups based on batterer background data (see Table 1 in the Appendix). Twenty-seven percent of the men in the control group considered themselves Hispanic compared to 19% for the treatment group. This variable was controlled for in the subsequent outcome models.

Overall, a typical subject in this sample was a male in his early thirties, with no official prior criminal history but high self-reported rates of psychological and physical abuse of the victim, with little more than a high school diploma, employed at least on a part-time basis, a household income of about $25,000 per year, and married to or living with his victim at the time of his arrest.

All the victims were females, with a mean age of thirty-two years.[2] Fifty-three percent reported being White, 21 percent reported being African American, 14 percent reported being some "other" race, 7 percent reported being a "mixed" race, 4 percent reported being American Indian, and 2% reported being Asian/Pacific Islander. In a separate question, 21% of the victims reported being of Hispanic origin. About 28 percent of the victims reported not having a high school degree, another 40 percent reported earning a high-school diploma or GED, and 32% reported completing some formal education beyond high school. Approximately half (54 percent) of the victims reported employment (either part-time or full-time) in the 30-day period before their arrest. The mean household income was approximately $20,000 per year, while the mean personal income level was about $10,000 per year. A majority of the victims reported having children (85 percent) and 58 percent of the victims had children with the batterer. No differences were observed between the treatment and control groups based on victim background data.

6. RESULTS

The results are presented in the following order: Batterer based 6-month follow-up measures of new reports of any domestic violence, physical abuse only, and controlling behavior (Tables 4 to 7 in the Appendix); victim based 6-month follow-up measures of new reports of any domestic violence, physical abuse only, and controlling behavior (Tables 8 to 11 in the Appendix); and new police arrests for any type of domestic violence within 6 and 12 months of the treatment/control condition assignment at the jail (Tables 12 to 13 in the Appendix). The first set of tables (Tables 4, 8 and 11 in the Appendix) for each of the three data sources (batterers, victims, and police arrests) contains descriptive/bivariate statistical comparisons, including the distribution of reported incidents of domestic violence by assigned treatment, and the percentage of zero to ten or more failures. Tables 4, 8 and 11include the mean, standard deviation and sample sizes for each outcome measure by assigned treatment. Also, a standardized effect size (this is the difference between the mean from the treatment group and the control group, divided by the pooled estimate of the standard deviation) is presented. The effect size statistic standardizes the difference between the experimental and control groups (i.e., the effect of the intervention) in terms of the standard deviation. Based on Cohen's (1988.) formulation a large effect size in the social sciences would be .40 (.40 versus .80), a medium effect size would be .25 (.40 versus .65), and a small effect size would be .10 (.40 versus .50). Tables 4, 8 and 11 include bivariate tests between the treatment and control group for each outcome measure. That is, chi-square testing is used for the zero failure compared to the one or greater failure comparison and ANOVA F test for comparing the mean score between the treatment and control groups. The purpose of Tables 4, 8 and 11 are to provide a clear

[2] Victim background data are only reported in the text and not in any of the tables.

presentation of the distribution of data for the treatment and control groups, prior to examining the data in the multivariate context. Typically, although not always, if there are not significant differences between the treatment and control groups in the simple bivariate context than there will not be differences in the multivariate context.

For each multivariate table (5, 6, 7, 9, 10, 11 and 13 in the Appendix) we present the results for the measures of prevalence, incidence/frequency, and time to first failure. The prevalence comparisons were conducted using logistic regression. The incidence/frequency of failure measures were examined using a count model (negative binomial regression) that tests whether the distribution of failures (i.e., frequency of cases in which new violence occurred) differed according to treatment/control assignment. The next set of analyses, using a type of hazard modeling called Cox regression, examines differences between treatment conditions in elapsed time to first failure. For each multivariate table (5, 6, 7, 9, 10, 11 and 13 in the Appendix) we present two types of models. The top rows in each of the multivariate tables present the Experimental Model, or our “Type 1" models. These initial models only specify the treatment assignment measure. The second set of rows present the Full Model, or our “Type 2" models. These final models specify a more complete set of structure measures/covariates, along with the treatment assignment measure. The full set of background covariates, includes: Age as a continuous variable (i.e., interval-level), race as a categorical variable, education as a continuous variable, employment as a dichotomous variable, personal income as a continuous variable, relation status as a categorical variable, prevalence of history of psychological abuse, prevalence of history of physical abuse, prior history of arrest for non-domestic violence offenses, and prior history of arrest for domestic violence offenses). The full model (Type 2) builds on this first model by adding additional control measures to account for the natural heterogeneity between and within the two comparison groups. These two models can be compared to see if the full model with covariates alters the significance of the treatment assignment variable.

The main purpose of the multivariate tables (5, 6, 7, 9, 10, 11 and 13 in the Appendix) is to see if the treatment assignment variable is statistically significant. For the logistic regression and negative binomial models, a statistically significant positive coefficient for the treatment variable means that the treatment group has a higher prevalence/frequency of recidivism than the control group. Conversely, for the logistic regression and negative binomial models, a statistically significant negative coefficient for the treatment variable means that the treatment group has a lower prevalence/frequency of recidivism than the control group (a sign that the program is having the desired effect). The time-to-failure model is based upon a Cox regression

that modeled the hazard of time-to-first new arrest within the first six and then twelve months of follow-up. With the Cox regression model a statistically significant negative coefficient for the treatment variable means that the treatment group had a percent reduction in the hazard rate on any given day after the treatment assignment (a sign that the program is having the desired effect). The opposite (i.e., an increase in the hazard rate) would be true if a statistically significant positive coefficient for the treatment variable emerged.

6.1 Batterer-based accounts of domestic violence recidivism results

As seen in Table 4 (in the Appendix), for both the treatment and control group the highest proportion of domestic violence is for psychological abuse. For both the treatment (46.2%) and control (43.4%) groups over 43% of the men in the whole sample admitted to psychological abuse, over 4% of the men in the whole sample admitted to threats of physical abuse (treatment 4.6% and control 5.7%), over 10% of the men in the whole sample admitted to physical abuse (treatment 10.8% and control 13.3%), and over 49% of the men in the whole sample admitted to at least one of these four types of domestic violence from the men in our sample (treatment 50.8% and control 49.1%). Also, the men in the whole sample admitted to on average over 2 acts of controlling behavior towards the victim (treatment 2 and control 3.2), the men in the whole sample admitted to on average over 7 acts of psychological abuse (treatment 7 and control 10.2), the men in the whole sample admitted to on average less than 1 act of threats of physical abuse (treatment 0.1 and control 0.5), the men in the whole sample admitted to on average to less than 1 act of physical abuse (treatment 0.4 and control 0.9), and the men in the whole sample admitted to on average over 9 acts of one of these four types of domestic violence (treatment 9.0 and control 14.8). Only one of the ANOVAs for the five batterer-based outcome comparisons between the treatment and control groups was statistically significant. The men in the treatment group (13.8%) had a lower prevalence rate of perpetrating controlling behavior than the men in the control group (30.2%) ($X2= 4.67$, $p<.01$). The ANOVA for the frequency of controlling behavior was not statistically significant (treatment mean=1.7 and control group mean=3.2), but (as will be seen later) this is due to the skewed nature of the data that requires a negative binomial regression to properly model the distribution of the data. The ANOVA results are presented for only descriptive purposes.

Table 5 (in the Appendix) contains batterer 6-month follow-up self-report data for any acts of domestic violence. The treatment variable in the type 1 and type 2 models for the logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure of batterer self-reported acts of any domestic violence. Given the absence of significant variables in all of the models on Table 5 (in the Appendix), all of the models had poor statistical fit (All R^2 were less than .15, and all p levels were non-significant).

Table 6 (in the Appendix) contains batterer 6-month follow-up self-report data for only physical abuse. The treatment variable in the type 1 and type 2 models for the logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure for batterer self-reported acts of physical abuse. All of the models in Table 6 had poor statistical fit (All R2 were less than .18, and all p levels were non-significant).

Table 7 (in the Appendix) contains batterer 6-month follow-up self-report data for only controlling behavior. The treatment variable in the type 1 and type 2 models for the logistic, negative binomial and Cox regression models were all statistically significant and in the predicted direction. The six-month Type 1 Prevalence Model shows that the treatment group had a lower proportion of participants using controlling behavior than the control group (type 1 model: b=-0.99, p<.05). The difference or effect size is also substantively large. The type 1 model estimated that 63 percent fewer than the expected number of subjects assigned to the treatment group committed a new incident of controlling behavior against the victim (ExpB= .372). The six-month Type 1 Frequency/Rate Model shows that the treatment group had a lower incidence or frequency of controlling behavior than the control group (type 1 model: b=-0.65, p<.001). The type 1 frequency model estimated that the treatment group committed 48 percent fewer than the expected frequency of incidents of controlling behavior against the victim (ExpB= .524). The six-month Type 1 Time-to-Failure Model shows that there was a significant decrease in the hazard rate for the treatment group compared with the control group (b = -1.33; p < .01). In other words, the men assigned to the treatment group had a 74 percent reduction in the likelihood of having a new incident of controlling behavior (ExpB= .264), thereby creating a longer period of non-controlling behavior for the victim.

All of the Type 2 models for controlling behavior (batterer self-reported) had similar results as the Type 1 models, with statistically significant treatment variables in the direction of showing the treatment to be effective. All of the Type 2 models in Table 7 (in the Appendix) had good statistical fit (All R^2 were nearly .30 and above, and there were a number of p levels that were statistically significant). Also, it is worth noting that one of the concerns raised earlier were that the men who completed the interviews were different than those who did not. Based on the earlier tests the author did not observe any differences and that when the saved selection process latent variables were used in each of the models in table 12 they each proved to be non-significant. These results suggest that there does not appear to be a selection process bias that could have determined the results from table 7.

6.2 Victim-based accounts of domestic violence recidivism results

As seen in Table 8 (in the Appendix), for both the treatment and control group the highest proportion of domestic violence is for controlling type behavior. For both the treatment (45.9%) and control (44.9%) groups about 45% of the sample of victims experienced controlling behavior from the men in our sample, over 60% of the sample of victims experienced psychological abuse from the men in our sample (treatment 63.5% and control 66.7%), over 23% of the sample of victims experienced threats of physical abuse from the men in our sample (treatment 25.7% and control 23.2%), over 30% of the sample of victims experienced physical abuse from the men in our sample (treatment 31.5% and control 31.9%), and over 67% of the sample of victims experienced at least one of these four types of domestic violence from the men in our sample (treatment 67.1% and control 72.5%). Also, the victims in this study experienced on average over 11 acts of controlling behavior from the men in our sample (treatment 13.8 and control 11.3), victims experienced on average over 21 acts of psychological abuse from the men in our sample (treatment 24.6 and control 21.9), victims experienced on average over 3 acts of threats of physical abuse from the men in our sample (treatment 3 and control 4.4), victims experienced on average over 11 acts of physical abuse from the men in our sample (treatment 12.7 and control 11.9), and victims experienced on average over 49 acts of one of these four types of domestic violence from the men in our sample (treatment 54.1 and control 49.7). None of the Chi-square tests or the ANOVAs for any of the five victim-based outcome comparisons between the treatment and control groups were statistically significant.

Table 9 (in the Appendix) contains victim 6-month follow-up self-report data for any acts of domestic violence. While there were a few background covariates (i.e., age, personal income, and history of psychological abuse) that were significant in Table 9, the treatment variable in the type 1 and type 2 models for the logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure of victim self-reported acts of domestic violence. All of the models in Table 9 (in the Appendix) had poor statistical fit (All R^2 were less than .25, and all p levels were non-significant).

Table 10 (in the Appendix) contains victim 6-month follow-up self-report data for only physical abuse. The treatment variable in the type 1 and type 2 models for the logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure for victim self-reported acts of physical abuse. All of the models in Table 10 had poor statistical fit (All R^2 were less than .18, and all p levels were non-significant).

Table 11 (in the Appendix) contains victim 6-month follow-up self-report data for only controlling behavior. In contrast to our finding in Table 7 (in the Appendix) suggesting the treatment program is reducing controlling behavior, the treatment variable in the type 1 and type 2 models for the logistic, negative binomial and Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6-month prevalence, frequency, and time-to-failure for victim reported acts of controlling behavior by the perpetrator.

6.3 Police domestic violence arrest results

Table 12 (in the Appendix) provides data on official police records of new domestic violence arrests at six and twelve months after the intervention. Neither the Chi-square tests nor the ANOVAs for any of the four arrest-based outcome comparisons between the treatment and control groups were statistically significant. The treatment (65.9%) and the control (65.7%) groups both had over 65% of the men in the sample re-arrested for a new domestic violence offense at six months post intervention (X^2=.002, p= NS). The treatment (mean=.90) and the control (mean= .88) groups both had on average less than 1 re-arrest for a new domestic violence offense at six months post intervention (F=.08, p=NS). The treatment (68.6%) and the control (69.6%) groups both had over 68% of the men in the sample re-arrested for a new domestic violence offense at twelve months post intervention (X^2=.07, p= NS). The treatment (mean=1.07) and the control (mean= 1.05) groups both had on average 1 re-arrest for a new domestic violence offense at six months post intervention (F=.06, p=NS).

Table 13 (in the Appendix) presents arrest data and includes not only 6-month recidivism data, but also 12-month recidivism data. The treatment variable in the type 1 and type 2 models for 6 and 12-month logistic regression models were all non-significant. That is, there was no difference between the treatment and control group in 6 and 12-month prevalence of new arrests. The treatment variable in the type 1 and type 2 models for 6 and 12-month negative binomial regression models were all non-significant. That is, there was no difference between the treatment and control group in 6 and 12-month frequency of new arrests. The treatment variable in the type 1 and type 2 models for 6 and 12-month Cox regression models were all non-significant. That is, there was no difference between the treatment and control group in 6 and 12-month time-to-failure of new arrests. The author also conducted a Cox Regression with a treatment by time-dependent interaction term to test for the proportionality of the hazard rates over the first 6 and 12 months of follow-up. The coefficient for the interaction term was non-significant in both the type 1 and type 2 models. Therefore, the interaction term was removed from the analysis to simplify the interpretation and presentation of the results. Given the absence

of significant variables in all of the models on Table 13, all of the models had poor statistical fit (All R^2 were less than .04, and all p levels were non-significant).

7. DISCUSSION AND CONCLUSION

Over the last few decades there have been many evaluations of batterer education programs. There are at least eight published literature reviews of over 40 single-site evaluations. However, most of these evaluations have lacked sufficient methodological rigor to yield conclusive answers about the effectiveness of batterer education programs. In fact, there have been only a handful of investigations that can make any legitimate claims about differences between treated and untreated batterers. In recent years, rigorous experimental evaluations have been completed on traditional court-mandated batterer education programs carried out in the community. Results from the Brooklyn experiment (Taylor et al., 2001) and the Palmer and colleagues (1992) study found that men assigned to batterer treatment had significantly lower rates of new incidents reported to criminal justice authorities than men assigned to the no treatment control group. However, no differences were observed between the treatment and control groups in these two studies on victim-based self-report measures of domestic violence recidivism. Also, the other true experiments by Dunford (2000) and Feder and Forde (2000) found no statistical differences between the treatment and control groups on new acts of domestic violence. With these mixed results for community based batterer treatment program in mind, we now have results for a somewhat different approach to batterer treatment.

While there has been a vast research literature on community-based programs for domestic violence (DV) offenders convicted of DV offenses, including experimental research, few programs have been developed for DV offenders awaiting trial. Also, little to no evaluation research has been conducted on these jail-based programs. The absence of these types of programs is worrisome, for many DV arrestees are able to pressure their victims into dropping the charges. In turn, these batterers do not get convicted and do not receive treatment. For many batterers, the time while they are held in jail after an arrest may be the only opportunity for a batterer intervention. This study evaluated the Sacramento Sheriff's jail-based batterer treatment program. The program was setup as an early intervention program to provide DV education for DV arrestees during their time of detention before going to court. The intervention was designed as a one-week batterer treatment program and was built on the idea that many of the men in the treatment program have never received a message about non-violence. The idea was that during the initial period of detention in the jail is a time when batterers might be receptive and (at the very least) be willing to listen to a message of non-violence. Gaining the attention of batterers in community-based treatment programs is rarely achieved (due to the numerous distractions available in the community that are not available in jail). The Sacramento Sheriff's jail-based batterer program attempts to capitalize on this window of opportunity and provide its

intervention. The hope is that by explicitly telling batterers that domestic violence is wrong and introducing them to non-violent approaches to handling conflict that a significant number of these men (of course not all of these men) will commit less domestic violence.

Another component of the program is to try and address some of the risk or aggravating factors for domestic violence. One of main risk factors for domestic violence is alcohol and drug use and this was explicitly addressed in the Sacramento program with drug and alcohol prevention education.

The current study is significant for it evaluates the effects of the Sacramento program on reducing repeat violence through one of the most rigorous designs available; that is, a true experimental design. As pointed out earlier, randomized experiments are typically thought of as the best method for eliminating threats to internal validity in evaluating social programs and provide the best counterfactual describing what would have happened to treatment group participants if they had not been exposed to the treatment (e.g., Berk et al. 1985). This research randomly assigned a large number of batterers (n=629) to either the batterer treatment wing of the jail or to a no-treatment control group in another wing of the jail. The implementation of the experiment was conducted without any "overrides" or interruptions to the random assignment process. Interviews were conducted with not only the victims (something done in most batterer treatment studies) but also the batterers shortly after the arrest and six months post-arrest. Additionally, official police arrest data on recidivism are analyzed for up to one year post-arrest.

Our main finding in this study relates to the batterer self reports on controlling behavior after six-months post intervention. The author observed for the prevalence model that over 60 percent fewer than the expected number of subjects assigned to the treatment group admitted to committing at least one new incident of controlling behavior against the victim (ExpB= .372). In terms of raw numbers, 13.8% of the treatment group participants admitted to a new act of controlling behavior compared to more than double the number of control group participants (30.2%). The frequency model estimated that the treatment group committed nearly 50 percent fewer than the expected frequency of incidents of controlling behavior within six months post intervention against the victim (ExpB= .524). In terms of raw numbers, on average 1.7 new acts of controlling behavior were admitted to by the treatment group participants compared to nearly double that number on average for the control group participants (3.2 acts of controlling

behavior).[3] The time-to-failure model revealed that there was a significant decrease in the controlling behavior hazard rate for the treatment group compared with the control group (over 70 percent reduction), thereby creating a longer period of time before a new controlling behavior act occurred. Also, in terms of raw numbers, more than 171 days on average expired before a new act of controlling behavior were admitted to by the treatment group participants (63 days for just those in treatment that failed) compared to only 145 days on average for the control group participants (48 days for just those that failed).[4]

Overall, what emerges from the analysis is that the batterer treatment program had an effect on the least serious form of domestic violence (controlling behavior prevalence, frequency and time-to-failure), based on the batterer self-reports (although no difference was observed based on the victim-based measure of controlling behavior). This is not to suggest that controlling behavior is not a serious problem. As discussed earlier, the controlling behavior subscale includes questions on keeping the victim from talking on the phone, spending time with friends, stopping her from going some place she wanted to go, stalking behavior, keeping her from using the family income, stealing money from her, taking away or threatening to take away her children, harming or threatening to harm her children, seriously harming or threatening to seriously harm other people she knows or suicidal type of behavior. Few would argue that this type of behavior does not represent serious violations of other's rights. There should not be much doubt that this behavior is harmful.

The fact that the Sacramento program was effective at reducing the prevalence, frequency and time-to-first failure for these problems is no doubt important to the victims who would have otherwise been on the receiving end of those acts. Which raises a question about whether there was an actual change in the controlling behavior of the treatment program participants, for the victim-based measure of controlling behavior found no differences between the treatment and control groups. In some ways it seems that the victim data on this issue would seem to be more revealing. The victim might have less of an interest in distorting the truth than the batterer – who might have a harder time acknowledging the harshness of his own behavior. On the other hand, the batterer treatment participants admitted to other types of more serious behavior like psychological and physical abuse at the same rate as the control group participants. If the

[3] While this comparison in the ANOVA model on Table 4 (in the Appendix) is not significant when properly modeled with a negative binomial regression in Table 7 (in the Appendix) it does appear statistically significant beyond the .001 level).

[4] These raw numbers are presented to give the reader a better sense of the distribution of the data. As with the frequency measure model, the appropriate time-to-failure Cox regression from Table 7 needs to be used to get an accurate assessment of the differences between the groups on this measure.

batterer treatment participants were lying why would they have been untruthful only about the controlling behavior and not the other more serious forms of violence? Also, the nature of some of the controlling behavior items might make it more likely for the batterers to have a more accurate assessment. For example, the batterer would be in the best position to know if they are stalking the victim. That is, the victim might not realize that the batterer has stopped stalking them or miss that the batterer has increased their stalking behavior. A similar case could be made for other items on the controlling behavior scale such as harming the victim's children or victim's friends, stealing money from the victim, or hiding family income.

Another possibility is that this finding is just an anomaly. However, this seems very unlikely due to the fact that all three aspects of this measure are significant (prevalence, frequency and time-to-failure).

Although not the focus of this study's intervention, one of the components of the Sacramento batterer treatment program was alcohol and drug education/prevention through a daily Twelve-Step Drug/Alcohol addiction support group. The program included these components because of their link to domestic violence perpetration. In the literature, lower drug and alcohol use is associated with less domestic violence (Copenhaver, 1998; Gondolf and Foster, 1991; Hamberger and Hastings, 1991; Kantor and Straus, 1990). Interestingly, two statistically significant differences were observed in batterer self-reported alcohol and marijuana use. Sixty percent of the treatment group consumed alcohol within the six-month period after the intervention compared to more than 77 percent for the control group (X^2=3.6, p<.05; overall n=118). Seventeen percent of the treatment group of batterers reported using any marijuana within the six-month period after the intervention compared to 30 percent for the control group (30%) (X^2=3.7, p<.05; overall n=118). While none of the other drug measures (related to other drugs, frequency or dependency) were statistically significant, it does suggest that the intervention was creating some type of changes in the treatment group.

While the data suggest that something has occurred with the men in the batterer treatment group, including changes in controlling behavior and less drug and alcohol use than the control group, for all the other measures the treatment program did not have an effect. This non-effect for all the other measures appears irrespective of the data sources (arrests, batterer/victim self-reports of domestic violence), type of measure (prevalence, frequency nor time-to-failure), and follow-up period (six or twelve months re-arrests). Also, the other two measures of psychological abuse and threats of physical abuse were also non-significant for both the Type 1 and Type 2 models for prevalence, frequency and time-to-failure (for the sake of parsimony these other non-significant models for victim and batterer reported surveys are not presented).

The author also looked at controlling behavior, psychological abuse, threats of physical abuse, and actual physical abuse with new partners the perpetrator was with in the six-month period after the triggering arrest. None of those measures were statistically significant for victim or batterer six month self report measures (these non-significant findings are also not presented).

Some cautionary statements

There are a number of potential and actual limitations to this study, including: The short duration of the intervention, the small sample size for the victim and batterer surveys, some inherent problems with arrest measures and self-reports, potential bias from low survey participation rates on substantive results, generalizeability of the results, and limited follow-up period.

Given the short duration of the intervention (for a week or less), it could be argued that the program developers were asking too much of the program to have an effect on all forms of domestic violence – including actual physical abuse. The one-week Sacramento jail program stands in opposition to the more typical length of 6 to 52 weeks for batterer treatment (Feazel et al., 1984; Healey and Smith, 1998) and some in the field have even advocated long-term treatment up to 5 years (Ewing, Lindsey, and Pomerantz, 1984). The program is restricted to a week or less in duration because that is about the maximum amount of time someone might be held after an arrest. However, the number of hours of batterer treatment were about three hours per day for the entire week – which meant over 20 hours for some men (which is not too much different than many of the 26 hour community-based batterer treatment programs. Also, the program included daily Twelve-Step Drug/Alcohol addiction support groups (something that is typically not done with the typical community-based batterer treatment programs), and strict regulations on television watching (special non-violent education programs were the only available programs). In some ways, the Sacramento jail program represented a type of intense impatient intervention where a lot is accomplished quickly compared to community-based outpatient programs that last much longer. Nevertheless, one possible enhancement would be for the Sacramento jail treatment group to refer cases to community-based programs to continue their treatment after they are released from jail. Also, in cases where the charges appear as if they will be dropped, some of these men could offered batterer treatment as part of a conditional discharge. In that way, the length of the intervention can be stretched.

While this study started out with a very large sample of 629 cases, only a small proportion of the victims (n=143) and batterers (n=119) in the sample participated in the surveys. With only 119 cases, the effect sizes would have needed to be in the medium (.25) sized range, based on Cohen's (1988) formulation, to find statistical significance (see Exhibit 1 below). For example, if the effect size were .22 (e.g., one group could be 50% and the other 28%) to have

power of 70% to yield a statistically significant result[5], .25 (e.g., one group could be 50% and the other 25%) to have power of 82% to yield a statistically significant result, and .28 (e.g., one group could be 50% and the other 22%) to have power of 90% to yield a statistically significant result. To find what Cohen (1988) refers to as a small effect size of .10 (.40 versus .50), with a sample size of 119 cases, to be statistically significant the power would only be 20%. The situation would be a bit better with the victim sample of 143 cases (see Exhibit 1 below). However, both the victim and batterer data sources offer much lower power levels than the much larger sample for the arrest database (which offers nearly 70% power to find even a .10 effect size). For the victim and batterer samples, this study was effectively not able to find small effect sizes as significant. However, given that the study had a very strong capacity to detect small effect sizes with the arrest data and was able to find effect sizes of .22 or greater at over 70% power for all the data sources this study is fairly strong in this area.

Exhibit 1: Effect size and Power Levels for Three Data Sources*

Power for Arrest data (n=582)	Power for Victims (n=143)	Power for Batterers (n=119)	Effect size
68%	23%	20%	.10 (Small)
96%	44%	38%	.15
100%	69%	61%	.20
100%	74%	66%	.21
100%	78%	70%	.22
100%	82%	74%	.23
100%	85%	78%	.24
100%	88%	82%	.25 (Medium)
100%	95%	90%	.28

* Calculated with the Power and Precision Software Version 1.20 (Borenstein, Rothstein and Cohen, 1997)

The measure with the most data available (only 7% or 47 of 629 cases were missing official record follow-up) was our arrest data. As mentioned above, the statistical power of the arrest data source was excellent. However, the data source where the study had its best statistical power is not without its own limitations. Arrest data are limited to those acts that are known to the police and have enough evidence to warrant an arrest. Crimes that never come to the attention of the police (the so called "dark figures" of crime) are therefore not included in our

[5] With a sample of 119 cases, the study has a 70% chance of avoiding a type II error. That is, this study will fail to detect a true difference of the desired size 30% of the time (on average). This means that if we ran this experiment 100 times, and if the true difference were of this size of .22, we would detect it as significant 70 times on average and fail to detect it 30 times on average.

arrest measure. Also, a number of the areas covered in our surveys are not considered crimes (e.g., some of the items on the controlling behavior and psychological abuse scales). Therefore, behavior that a person could not be arrested for is not something that is picked up by our arrest measure. While one of the strengths of the arrest measures are that they are not dependent on a person's memory, are not subject to underreporting or overreporting, are fairly inexpensive to collect for both six and twelve month follow-up, it is not available for measures such as controlling behavior. This is unfortunate due to the mixed findings between the batterer and victim accounts on the controlling behavior measure.

More generally arrest measures are not effective if batterers learn to abuse the victim in such a way that they are less likely to call the police. Arrest measures are highly dependent on a victim's willingness to file a complaint or call the police. This raises the possibility that assignment to treatment versus a control group may differentially affect the victim's willingness to contact criminal justice officials when future abuse occurs (i.e., a possible instrumentation effect). A victim may not report her partner's abuse for a number of reasons. This includes the possibility that she might prefer to see her partner continue in treatment where she believes it will eventually lead to changes in his abusive behavior rather than take the risk of reporting his continued abuse and see him go to jail. Alternately, a victim may resent the criminal justice system's intrusion into her life in the form of mandating a treatment that she is then responsible for paying. Most programs require the abuser to pay for the treatment and by extension that means that it is the family that pays for the treatment (Zorza, 2003). If the victim views the treatment as ineffective, it may make her critical and suspicious of the police and less likely to cooperate in the case of reporting future incidences of abuse. The dependence of arrest reports on the behavior of the victim allows for the plausibility that different rates between batterers in the treatment and control groups may reflect a measurement artifact and not a genuine treatment effect.

The next concern is whether our sample of victim and batterers is biased due to the low survey participation rates. Low response rates are a problem because the cases in which follow-up data are available may be different than those cases in which data are not available. For example, Edleson and Syers (1990) reported higher levels of education and income for batterers who completed follow-up surveys than for batterers who did not. It is unclear; therefore, whether their analysis of treatment effects applies to the low SES batterers who failed to complete the survey as well as the higher SES batterers who did. Most batterer treatment studies suffer from low survey participation rates. The survey participation rates for victim surveys was 70% for the Dunford (2000b) study, 50% for the Taylor et al. (2001) study, about 20% for the Feder and Dugan (2002) study and 41% for the Harrell (1991) study. Low survey participation rates could be a problem in light of research indicating that certain victims of domestic violence are more likely to be lost in the research follow-up than are others. This

research strongly suggests that women victims of domestic violence who are more difficult to retain in follow-up research are both more marginal and more likely to be more frequently and severely abused (Sullivan, Rumptz, Campbell, Eby and Davidson, 1996).

The batterer six-month self-reports were available for only about 20% of the sample. This stands in contrast to the batterer baseline completion rate of 90% (due to the availability of the men for interviewing while in jail at the time of the baseline period). The victim six-month follow-up self-reports were available for only about 23% of the sample (down from a 33% participation rate for the victim baseline interviews). Essentially, the concern is that the achieved sample for this study may not represent all of the treatment cases and perhaps be slanted towards certain type of cases that are more or less resilient to change than the average case (Edleson, 1996; Gondolf, 1997). First, this might simply mean that the study results only apply to a more limited group of people (e.g., the results would not apply to those unemployed if the study team were not able to interview any unemployed persons). However, the situation would be much worse if the attrition rate were different between the treatment and control groups and/or the characteristics of the non-completers were different between the treatment and control groups. That is, if one of the groups had higher attrition among the cases with a higher risk to recidivate than that would bias the test against that group. To investigate the impact of attrition, the author conducted a number of analyses and found no measurable differences between the victim and batterer survey participants and non-participants (on the arrest and baseline batterer interview measures that existed for nearly all the cases) as a whole. Also, no systematic patterns emerged for our analyses between the victim and batterer completers to non-completers across the treatment and control groups. The best indications are that the attrition issue did not have a measurable substantive impact on the results of this study.

One of the areas that the Taylor et al. (2001) study was criticized for was the generalizeability of the results, since the sample represented a small proportion of spouse abuse cases that the Brooklyn Criminal Court adjudicated during the intake period (376 cases out of nearly 11,000 adjudicated cases met their study criteria). Similarly, many of the other evaluations of batterer treatment programs can be criticized on the grounds that participants represent only a small and probably unrepresentative proportion of batterers processed in the courts from which their samples were drawn. In the current study data were collected for about 10 months on 629 arrest cases for male perpetrated domestic violence in Sacramento County. During that same period, the Sacramento County Main Jail booked approximately 50,000 male arrestees. Of those 50,000 male arrestees, 5% or 2,500 of those were arrested for DV offenses (with about 7 DV arrestees per day). The batterer program excluded 50% of the available pool

of DV arrestees because of prior convictions. Therefore, the program's eligibility pool was 1,250 cases for the year (50% of 2,500 DV arrestees) or 1,040 cases for the 10-month study intake period. From the 1,040 cases the study lost an additional 370 cases due to the arrestees making bail and 41 cases for a variety of other reasons (e.g., arrestees deemed to dangerous to be interviewed for a research project). Nevertheless all persons that were eligible for the study were entered into the experiment, as opposed to the Taylor et al. study (2001) where the prosecutor, defendant and judge had to agree to treatment in order for a case to be eligible for the random assignment process. Also, this study included both motivated and unmotivated arrestees in the sample; people could not choose to opt out of the experiment (while they could refuse to participate in the survey, the study team still had official record data on nearly the entire sample).

Finally, the follow-up period was limited to only six months for victim and batterer self-report and only 12 months for arrest measures. While our main finding related to controlling behavior was significant at the 6-month follow-up period, the question about whether this finding would holdup over a longer 12-month follow-up is still open. Additionally, funding was limited for this project, but a longer arrest follow-up beyond 12 months could be carried out fairly easily. Also, the official measures were restricted to arrests due to funding limitations, but the study might have also included incidents or complaints that the police record that do not lead to an arrest.

Concluding statement

Studies of batterer treatment programs have grown increasingly sophisticated, using designs with a high degree of internal validity. As recognized by Fagan (1996), randomized experiments are the preferred evaluation designs. The few experimental evaluations that have been completed provide useful information on statistical power and sample sizes for those designing future studies. As studies are completed, we may know with some confidence how much and under what conditions treatment programs reduce violence. Then attention will turn to policy issues of how these programs are conducted and suggested revisions like more formal links between jailed based and community based batterer treatment programs can be established. Also, there is disagreement in the field over the content of treatment sessions. Many programs have adopted a feminist orientation that assumes that eliminating violent behavior requires changing the participant's perception of the roles of men and women in society. Other programs take a more pragmatic approach, emphasizing anger control, stress management, and better communication techniques. There currently is no empirical basis to decide which approach is best.

To date, this study has provided one of the most rigorous tests of the effectiveness of jail-based batterer treatment. The experiment was implemented well. There were no overrides in

this study (i.e., where the assigned condition to treatment or control group is bypassed in favor of something else). This study addressed some methodological problems found in prior research, such as disentangling the effects of treatment from sample selection effects, including a large sample of treatment and control group cases, studying the total population to be evaluated (as opposed to only those who completed the program), and proper specification of the primary outcome measures of domestic violence using reliable measures. Also, a number of tests were conducted that suggest that any of the potential biases that might have been introduced into this study did not have a measurable effect on the substantive results of this study. Aside from the possible civil liberty objections to the program (i.e., people are being given treatment prior to any finding of guilt), the program is not making things worse for the victims and some of the evidence suggests that it might in fact be doing some good in reducing controlling behavior and alcohol and drug use.

8. REFERENCES

Adams, D. and McCormick, A. (1982). Men unlearning violence: A group approach based on the collective model. In M. Roy (Ed.), The abusive partner: An analysis of domestic battering (pp. 170-197). New York: Van Nostrand Reinhold.

Armitage, P. (1996). The design and analysis for clinical trials. In S. Ghosh and C. R. Rao (Eds.), *Design and Analysis of Experiments*. Handbook of Statistics, vol. 13. Amsterdam: Elsevier.

Babcock, J.C., Green, C.E. and C. Robie (In press). Does Batterers' Treatment Work?: A Meta-Analytic Review of Domestic Violence Treatment. Clinical Psychology Review.

Babcock, J. C. and Taillade, J. (2000). Evaluating interventions for men who batter. In J. Vincent and E. Jouriles (Eds.), Domestic violence: Guidelines for research-informed practice (pp.37-77). Philidelphia: Jessica Kingsley.

Bachman, Ronet, Raymond Paternoster, and Sally Ward. "The Rationality of Sexual Offending: Testing a Deterrence/Rational Choice Conception of Sexual Assault." *Law and Society Review* 26, no. 2 (1992): 343-75.

Berk, R. A., Black, H., Lilly, J., and Rikoski, G. (1991). *Evaluating Alternative Police Response to Spouse Assault in Colorado Springs: An enhanced replication of the Minneapolis experiment, 1987-1989* [Final Report]. Colorado Springs, CO: Colorado Springs Police Department.

Berk, R. A., Smyth, G. K., and Sherman, L. W. (1988). When random assignment fails: Some lessons from the Minneapolis Spouse Abuse Experiment. *Journal of Quantitative Criminology*, *4*(3), 209-233.

Blumstein, A., Cohen, J., Roth, J., and Visher, C. (Eds). (1986). *Criminal Careers and "Career Criminals."* Washington, D.C.: National Academy ~~of~~ Press.

Borenstein, M., Rothstein, H. and J. Cohen, (1997). Power and Precision Software Version 1.20. New Brunswick, New Jersey.

Brisson, N. (1981). Battering husbands: A survey of abusive men. Victimology, 6, 338-344.

Buzawa, E. S., and Buzawa, C. G. (1996). *Domestic violence: The criminal justice response* (2nd). Newbury Park, CA: Sage Publications.

Buzawa, E. and Buzawa, C. (1996). Domestic violence: The criminal justice response (2nd edition). Thousand Oaks, CA: Sage.

Campbell, D. T., and Stanley, J. C. Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally, 1966.

Carmody, Diane C., and Kirk R. Williams. "Wife Assault and Perceptions of Sanctions." *Violence and Victims* 2, no. 1 (1987): 25-38.

Chen, H. T., Bersani, C., Myers, S., and Denton, R. (1989). Evaluating the effectiveness of a court sponsored abuser treatment program. *Journal of Family Violence, 4*(4), 309-322.

Chen, H.C., Bersani, S.C., and Denton, R. (1989). Evaluating the effectiveness of a court-sponsored abuser treatment program. Journal of Family Violence, 4 , 309-322.

Cohen J (1992). A power primer. Psychological Bulletin, 112, 155-159
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd edition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen J (1983). Statistical power analysis for the behavioral sciences (2nd ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen J (1965). Some statistical issues in psychological research. In BB Wolman (Ed), Handbook of clinical psychology (pp 95-121). New York: McGraw Hill.

Copenhaver, M.M.(1998). Testing A Social-Cognitive Model of Intimate Abusiveness Among Substance Dependent Males. Virginia Polytechnic Institute, Unpublished dissertation.

Crowell, N. A., and Burgess, A. W. (Editors). (1996). *Understanding violence against women*. Washington, D.C.: National Academy Press.

Daniels, J.W. and Murphy, C.M. (1997). Stages and processes of change in batterers' treatment. Cognitive and Behavioral Practice, 4, 123-145.

Davis, R. C., and Taylor, B. G. (1997). A proactive response to family violence: The results of a randomized experiment. *Criminology*, *35*(2), 307-333.

Davis, R.C., and Taylor, B.G. (1999). Does batterer treatment reduce violence? A synthesis of the literature. Women and Criminal Justice, 10 (2) 69-93.

Davis, R. C., and Taylor, B. G. (1999). Does batterer treatment reduce violence? A synthesis of the literature. *Women and Criminal Justice*, *10*(2), 69-93.

Davis, R.~~,~~ C., Smith, B. E., and Nickles, L. (1997). *Prosecuting domestic violence cases with reluctant victims: Assessing two novel approaches*. Washington, D.C.: American Bar Association.

De Leon, George, and Nancy Jainchill. "Circumstances, Motivation, Readiness and Suitability." *Journal of Psychoactive Drugs* 18 (1986): 203-08.

Dubin, J. A., and Rivers, D. (1989, November). Selection bias in linear regression, logit and probit model. *Sociological Methods and Research*, 360-390.

Dutton, D. G. (1988). Profiling of wife assaulters: Preliminary evidence for a trimodal analysis. *Violence and Victims*, *3*(3), 5-29.

Dutton, D. (1988). Research advances in the study of wife assault: Etiology and prevention. Law and Mental Health, 4, 161-220.

Dutton, D. (1986). The outcome of court-mandated treatment for wife assault: A quasi-experimental evaluation. *Violence and Victims*, *1*(3), 163-175.

Dobash, R. P., Dobash, R .E., Cavanagh, K., and Lewis, R. (1995). Evaluating criminal justice programmes for violent men. In R. E. Dobash, R. P. Dobash and L. Noaks (Eds.), Gender and crime. Cardiff, Wales: University of Wales Press.

Dobash, R., Dobash, R .E., Cavanagh, K., and Lewis, R. (1996). Re-education programmes for violent men--an evaluation. Research Fundings, 46, 1-4.

Dobash, R., Dobash, R., Cavenagh, K., and Lewis, R. (1996). *Research evaluation of programmes for violent men*. Edinburgh, Scotland: Centre Research, Scottish Office (141).

Dunford, F.W. (2000). The San Diego Navy Experiment: An assessment of interventions for men who assault their wives. Journal of Consulting and Clinical Psychology, 68 , 468-476.

Dutton, D.G. (1986). The outcome of court mandated treatment for wife assault: A quasi-experimental evaluation. Violence and Victims, 1 , 163-175.

Edleson, J. (1996). Controversy and change in batterers' programs. In J. Edleson and Z.

Edleson, J., and Syers, M. (1990). Relative effectiveness of group treatments for men who batter. *National Association of Social Workers, Inc.*, 10-17.

Eisikovits (Eds.), Future interventions with battered women and their families. Thousand Oaks, CA: Sage Publications, 154-169.

Eisikovits, Z., and Edleson, J. (1989, September). Intervening with men who batter: A critical review of the literature. *Social Service Review*, 384-414.

Ewing, W., Lindsey, M. and Pomerantz, J. (1984). Battering: An AMMEND manual for helpers. Denver, CO: AMMEND.

Fagan, J. A., and Browne, A. (1994). Violence against spouses and intimates. In A. J. Reiss Jr. and J. A. Roth (Eds.), *Understanding and controlling violence* (Vol. 3). Washington, D.C.: National Academy Press.

Fagan, J. A. (1996). The criminalization of domestic violence: Promises and limits. Conference on Criminal Justice Research and Evaluation. Washington, D.C.: National Institute of Justice. need a year on the conference

Fagan, J. A., Garner, J., and Maxwell, C. D. (1997). *Reducing injuries to women in domestic assaults* [Final Report]. Centers for Disease Control and Prevention, Department of Public Health and Human Services.

Fagan, J., Friedman, E., Wexler, S., and 'Lewis, V. L. (1984). *National Family Violence Evaluation* [Final Report]. San Francisco: URSA spell out Institute.

Farley, D. and Magill, J. (1988). An evaluation of a group program for men who batter. Social Work With Groups, 11(3), 53-65.

Fawcett, G., Heise, L.L. Espejel, L. and Pick, S. (1999). Changing community responses to wife abuse: A research and demonstration project in Iztacalco, Mexico. American Psychologist, 54, 41-49.

Feazell, C.S., Mayers, R.S., and Deschner, J. (1984). Services for men who batter: Implications for programs and policies. Family Relations, 33, 217-223.

Feder, L., and Forde, D.R. (2000). A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward experiment . National Institute of Justice.

Feder, L. (1996). *A test of the efficacy of court-mandated counseli9ng for domestic violence: A Broward County experiment* [Proposal submitted to the National Institute of Justice]. Boca Raton, FL: Florida Atlantic.

Feder, L. (1999). The efficacy of court-mandated counseling for convicted misdemeanor domestic violence offenders: The Broward experiment. American Society of Criminology. Washington, D.C. publication or press?

Garner, J., Fagan, J. A., and Maxwell, C. D. (1995). Published findings from the spouse assault replication program: A critical review. *Journal of Quantitative Criminology*, *11*(1), 3-28.

Gartin, P. R. (1995). Dealing with design failures in randomized field experiments: Analytic issues regarding the evaluation of treatment effects. *Journal of Research in Crime and Delinquency*, 425-445.

Gelber, R. D., and Zelen, M. (1986). Planning and reporting of clinical trials. In P. Calabresi, P. S. Schein and S. A. Rosenberg (Eds.), *Medical oncology* (406-425). New York, NY: Macmillian Publishing Company.

Goolkasian, G. (1986). Confronting domestic violence: The role of criminal court judges.

Washington, DC: National Institute of Justice.

Gondolf, E. (1987). Evaluating programs for men who batter: Problems and prospects. Journal of Family Violence, 2(1), 95-108.

Gondolf, E. (1995). *Batterer intervention: What we know and need to know.* National Institute of Justice Violence Against Women Strategic Planning Meeting. Washington, D.C. need a year

Gondolf, E. (1997). Expanding batterer program evaluations. In G. K. Kaufman and J. Jasinski (Eds.), *Out of darkness: Contemporary research perspectives on family violence.* Thousand Oaks, CA; Sage Publications, 208-218.

Gondolf, E.W. (1999a). A comparison of four batterer intervention systems: Do court referral, program length, and services matter? Journal of Interpersonal Violence, 14 , 41-61.

Gondolf, E.W. (1999b). MCMI-III results for batterer program participants in four cities: Less pathological than expected. Journal of Family Violence, 14 , 1-17.

Gondolf, E.W., and Foster, R.A. (1991). Wife assault among VA alcohol rehabilitation patients. Hospital and Community Psychiatry, 42, 74-79.

Grasmick, Harold G., and Robert J. Jr. Bursik. "Conscience, Significant Others, and Rational Choice: Extending the Deterrence Model." *Law and Society Review* 24, no. 3 (1990): 837-61.

Greenfeld, L. A., Rand, M. R., Craven, D., Klaus, P., Perkins, C. A., Ringel, C., Warchol, G.,

Maston, C., and Fox, J. A. (1998). *Violence by intimates: Analysis of data on crimes by current or former spouses, boyfriends, and girlfriends*. Bureau of Justice Statistics Factbook, vol. NCJ-167237. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs (46).

Hamberger, L. K. and Hastings, J. (1989). Counseling male spouse abusers: Characteristics of treamtent completers and dropouts. Violence and Victims, 4(1), 275-286.

Hamberger, L.K., and Hastings, J.E. (1991). Personality correlates of men who batter and nonviolent men: Some continuities and discontinuities. Journal of Family Violence, 6, 131-147.

Hamberger, K., and Hastings, J. (1993). Court-mandated treatment of men who assault their partner: Issues, controversies and outcomes. In Z. Hilton (Ed.), Legal responses to wife assault: Current trends and evaluation. Newbury Park, CA: Sage Publications, 188-232

Healey, K., Smith, C., and O'Sullivan, C. (1998). Batterer intervention: Program approaches and criminal justice strategies. Washington, DC: Department of Justice.

Hanna, C. (1996). No right to choose: Mandated victim participation in domestic violence prosecutions. Harvard Law Review, 109(8), 1849-1910.

Harrell, A.V. (1991). Evaluation of court ordered treatment for domestic violence offenders . Final report submitted to the State Justice Institute. Washington, DC: The Urban Institute.

Healey, K.M. and Smith, C. (1998). Batterer programs: What criminal justice agencies need to know. Research in Brief. NCJ 171683. National Institute of Justice: Washington, DC.

Healey, K., Smith, C., and O'Sullivan, C. (1998). *Batterer intervention: Program approaches and criminal justice strategies*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 153-161.

Hilberman, E. (1980). Overview: The "wife-beater's wife" reconsidered. American Journal of Psychiatry, 137(11), 1336-1347.

Hirschel, J. D. and Hutchinson, I. (1992). Female spouse abuse and the police response: The Charlotte, North Carolina Experiment. Journal of Criminal Law and Criminology, 83(1), 73-119.

Hotaling, G. T., and Sugarman, D. (1986). An analysis of risk markers in husband to wife violence: The current state of knowledge. *Violence and* v*ictims, 1*(2), 101-124.

Jennings, J. (1987). History and issues in the treatment of battering men: A case for unstructured group therapy. Journal of Family Violence, 2(3), 193-213.

Johnson, J. and Kanzler, D. (1993). Treating domestic violence: Evaluating the effectiveness of a domestic violence diversion program. Studies in Symbolic Interaction, 15, 271-289.

Land, Kenneth, C., Patricia L. McCall, and Daniel S. Nagin. "A Comparison of Poisson, Negative Binomial, and Semiparametic Mixed Regression Models." Sociological Methods and Research 24, no. 4 (May 1996): 387-442.

Langan, P. and Innes, C. (1986). Preventing domestic violence against women. Washington, DC: National Institute of Justice.

Lind, E. Allan, Ruth Kanfer, and P. Christopher Early. "Voice, Control and Procedural Justice: Instrumental and Noninstrumental Concerns in Fairness Judgements." *Journal of Personality and Social Psychology* 59 (1990): 952-59.

Little, R. J. A., and Schenker, N. (1995). Missing data. In G. Arminger, C. C. Clogg and M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavior science*. New York, NY: Plenum Press, 39-76.

Manning, Peter K. "The Preventive Conceit: The Black Box in Market Context." *American Behavioral Scientist* 36, no. 5 (May 1993): 639-50.

Mason, Robert L., Richard F. Gunst, and James L. Hess. Statistical Design and Analysis of Experiments: With Applications to Engineering and Science. Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statisticts. New York, NY: John Wiley and Sons, Inc., 1989.

Maxwell, C. D. (1998). *The specific deterrent effect of arrest on aggression between intimates and spouses* [dissertation], Newark, NJ: Rutgers, the State University of New Jersey.

Maxwell, S. R. (1994). Formal and informal social controls of drug treatment retention of offenders [dissertation]. Newark, NJ: Rutgers, the State University of New Jersey.

Miller, T., Cohen, M., and Wiersema, B. (1996). Victim costs and consequences: A new look. Washington, DC: National Institute of Justice.

Murphy, C.M. and Baxter, V.A. (1997). Motivating batterers to change in the treatment context. Journal of Interpersonal Violence, 12, 607-619.

Nagin, Daniel S., and Raymond Paternoster. "Personal Capital and Social Control: The Deterrence Implications of a Theory of Individual Differences in Criminal Offending." *Criminology* 32, no. 4 (1994): 581-.

Kantor, G.K., and Straus, M.A. (1990). The "drunken bum" theory of wife beating. In M.A. Straus and R.J. Gelles (Eds.), Physical violence in American families (pp. 203-224). New Brunswick, NJ: Transaction.

Palmer, S.E., Brown, R.A., and Barrera, M.E. (1992). Group treatment program for abusive husbands: Long term evaluation. American Journal of Orthopsychiatry, 62 , 276-283.

Patel, H. I. (1996). Clinical trials in drug development: Some statistical issues. In S. Ghosh and C. R. Rao (Eds.), *Design and analysis of experiments*. Handbook of Statistics, vol. 13. Amsterdam: Elsevier.

Paternoster, Raymond. "The Deterrent Effect of the Perceived Certainty and Severity of Punishment: A Review of the Evidence and Issues." *Justice Quarterly* 4, no. 2 (1987): 173-217.

Paternoster, Raymond, and Robert Brame. "Mutiple Routes to Delinquency? a Test of Developmental and General Theories of Crime." *Criminology* 35, no. 1 (February 1997): 49-84.

Pirog-Good, M. and Stets-Kealey, J. (1985). Male batterers and battering prevention programs: A national survey. Response, 8, 8-12.

Rebovich, D. J. (1996). Prosecution response to domestic violence: Results of a survey of large jurisdictions. In E. S. Buzawa and C. G. Buzawa (Eds.), Do arrests and restraining orders work? Thousand Oaks, CA: Sage.

Rennison, C. R. and Welchans, S. (2000). Intimate partner violence. Washington, DC: National Institute of Justice.

Rosenbaum, A., and O'Leary, K. D. (1986). The treatment of marital violence. In N. S. Jacoben and A. S. Gurmand (Eds.), *Clinical handbook of marital therapy*. The Guilford family therapy series.

Rosenfield, B. D. (1992). Court ordered treatment of spouse abuse. *Clinical Psychological Review, 12*, 205-226.

Saunders, D. G., and Azar, S. T. (1989). Treatment programs for family violence. In L. Ohlin and M. Tonry (Eds.), *Family Violence*. Crime and justice: a review of research, vol. 11. Chicago, IL: The University of Chicago Press.

Saunders, D. (1996). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and victims, 11*(4), 393-414.

Sherman, L. W. (1992a). *Policing domestic violence: Experiments and dilemmas*. New York: Free Press.

Sherman and Smith 1992 Sherman, Lawrence W., et al. "From Initial Deterrence to Long-Term Escalation: Short Custody Arrest for Poverty Ghetto Domestic Violence." Criminology 29, no. 4 (1991): 821-50.

Snyder, D. and Scheer, N. (1981). Predicting disposition following brief residence at a shelter for battered women. American Journal of Community Psychology, 9, 559-566.

Straus, M. A. (1979). Measuring family conflict and violence: The conflict tactics scale. *Journal of Marriage and the Family*, *41*, 75-88.

Sullivan, C., Rumptz, M., Campbell, R., Eby, K., & Davidson, W. (1996). Retaining participants in longitudinal community research: A comprehensive protocol. Journal of Applied Behavioral Science, 32(3), 262-276.

Taylor, B.G., Davis, R.C., and Maxwell, C.D. (2001). The effects of a group batterer treatment program in Brooklyn. Justice Quarterly, 18 , 170-201.

Tjaden, P. and Thoennes, N. (1998). Prevalence, Incidence, and Consequences of Violence Against Women: Findings From the National Violence Against Women Survey. Research in Brief. National Institute of Justice, Government printing office, Washington, DC.

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

Tolman, R. M., and Edelson, J. L. (1995). Interventions for men who batter: A review of research. In S. M. Stith and M. A. Straus (Eds.), *Understanding partner violence: Prevalence, causes, consequences, and solutions*. Minneapolis, MN: National Council on Family Relations.

Tyler, Tom R. *Why People Obey the Law*. New Haven and London: Yale University Press, 1990.

Weisberg, S. (1985). *Applied linear regression*. New York, NY: John Wiley and Sons.

Widom, C. S. (1992). The cycle of violence. Washington, DC: US Department of Justice.

Williams, Kirk R. "Social Sources of Marital Violence and Deterrence: Testing and Integrated Theory of Assaults Between Partners." *Journal of Marriage and the Family* 54 (August 1992): 620-29.

Williams, Kirk.R., and Richard. Hawkins. "Perceptual Research on General Deterrence: A Critical Review." *Law and Society Review* 20 (1986): 545-72.

Williams, Kirk R., and Richard Hawkins. "Controlling Male Aggression in Intimate Relationships." *Law and Society Review* 24, no. 4 (1989a): 591-612.

Williams, Kirk R., and Richard Hawkins. "The Meaning of Arrest for Wife Assault." *Criminology* 27, no. 1 (1989b): 163-81.

Williams, Kirk R., and Richard Hawkins. "Wife Assault, Cost of Arrest, and the Deterrence Process." *Journal of Research and Crime Delinquency* 29 (1992): 292-310.

Winship, C., and Mare, R. D. (1992). Models for sample selection bias. In *Annual Review of Sociology*. Palo Alto, CA: Annual Reviews Inc., 327-350.

Zorza, J. (2003). New research: Broward County Experiment shows no benefit from batterer intervention programs. Domestic Violence Report, 8, 23-25.

Table 1. Batterer's Characteristics by Assigned Treatment

	Total	Treatment	Control
Total N of Assigned Batterers	629	317	312
Age (mean)	32.8	32.0	33.6
Race/Ethnicity			
White	34.3%	33.8%	34.7%
African American	28.3%	29.6%	27.0%
Asian or pacific islander	3.0%	3.5%	2.5%
American Indian	3.9%	3.2%	4.6%
Mixed race	13.2%	13.0%	13.3%
Other race	17.4%	16.9%	17.9%
Percent of hispanic origin	23.2%	19.0%	27.4%*
Education			
No Diploma/GED	34.2%	35.7%	32.6%
High School Diploma/GED	38.9%	36.4%	41.4%
More than High School	26.9%	27.9%	26.0%
Employed in the 30 days before the arrest	74.2%	72.5%	75.8%
Household income (mean)	$ 25,706	$ 25,996	$ 25,433
Personal Income (mean)	$ 17,851	$ 18,137	$ 17,573
Relationship Status			
Married	37.4%	35.2%	39.6%
Separated/divorced	1.3%	1.1%	1.5%
Living together	39.8%	40.4%	39.3%
Intimate not living together	9.6%	9.6%	9.6%
Past intimate	11.9%	13.7%	10.0%
Prevalence of domestic violence ever before arrest (self-reported by batterers)			
Psychological abuse	88.0%	89.8%	86.0%
Physical abuse	54.3%	53.6%	55.0%
Prior Arrest for non-domestic violence	24.4%	25.8%	23.0%
Prior Arrest for domestic violence	36.1%	34.4%	37.8%

Note: All reported data (except for the data on prior arrest) are from the baseline interview with the batterers.

*This is the only significant difference across all of the comparisons ($X2 = 5.6$, $p<.05$)

Table 2. Implementation of Experimental Design

	Treatment less than a full day vs full day or more		Six Month Follow-up Interviews			
			Victim		Batterer	
Base Rate	29%	n=91	23%	n=143	19%	n=119
Logistic Regression	b	Exp(B)	b	Exp(B)	b	Exp(B)
Age	0.04	1.04 *	0.03	1.03	0.03	1.03
Race/Ethnicity (White reference)		ns	ns			
African American	-0.50	0.60	-0.61	0.54	-0.49	0.61
Asian or pacific islander	-1.06	0.35	0.57	1.77	-6.09	0.01
American Indian	0.09	1.09	0.07	1.07	0.29	1.33
Mixed race	0.34	1.40	0.13	1.13	-0.61	0.54
Other race	0.02	1.02	-0.86	0.42	-0.04	0.96
Education - highest completed grade	0.01	1.01	0.19 *	1.22	0.01	1.01
Employed in the 30 days before the arrest (yes=1)	-0.17	0.84	0.74	2.09	0.30	1.35
Personal Income (mean)	0.01	1.00	0.01	1.00	0.02	1.01
Relationship Status (Married reference)		ns	ns			
Separated/divorced	0.94	2.57	3.84	4.60	0.85	2.33
Living together	0.62	1.87	-0.33	0.72	-0.17	0.84
Intimate not living together	-1.14	0.32	-1.17	0.31	-0.18	0.84
Past intimate	0.34	1.41	0.21	1.24	0.31	1.37
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	-0.17	0.85	0.07	1.08	0.12	1.13
Prevalence of physical abuse ever before arrest (self-reported by batterers)	0.29	1.34 **	0.04	1.04	0.04	1.04
Prior Arrest for non-domestic violence (yes=1)	-0.13	0.88	-0.09	0.92	0.02	1.02
Prior Arrest for domestic violence (yes=1)	0.05	1.05	0.11	1.11	0.06	1.06
Treatment Assigned (1= treatment; 2= control)			-0.01	0.99	-0.39	0.68
Intercept	-2.30	ns	-3.28 *	0.04	-2.36	0.09
Cox & Snell pseudo R^2	0.10		0.10		0.10	
Final Log likelihood	258.11		267.11		457.17	
P-value	0.07		0.02		0.01	

* = p-value < 0.05; ** = p-value < 0.01; *** = p-value < 0.001

Table 3: Chronbach's Alpha Scores

		Victim 6 month followup	Batterer 6 month followup
Controlling behavior scale (11 items)	Prevalence	0.79	0.64
	Frequency	0.79	0.70
Psychological abuse scale (8 items)	Prevalence	0.82	0.72
	Frequency	0.84	0.80
Threats with physical abuse (5 items)	Prevalence	0.70	0.65
	Frequency	0.70	0.20
Physical abuse scale (15 items)	Prevalence	0.91	0.72
	Frequency	0.91	0.64

Table 4. Frequency of batterer reported incidents by assigned treatment: Six month follow-up interview data

	Any DV		Controlling		Psychological Abuse		Threats of Physical Abuse		Physical Abuse	
	Control %	TXT %	Control %	TXT %	Control %	TXT %	Control %	TXT %	Control %	TXT %
0	50.9%	49.2%	69.8%	86.2%	56.6%	53.8%	94.3%	95.4%	86.7%	89.2%
1	1.9%	6.2%	5.7%	4.6%	0.0%	3.1%	0.0%	3.1%	5.7%	1.5%
2	1.9%	3.1%	7.5%	1.6%	3.8%	3.1%	1.9%	0.0%	1.9%	3.3%
3	3.7%	4.6%	7.5%	0.0%	1.9%	4.6%	1.9%	1.5%	0.0%	1.5%
4	3.8%	3.1%	0.0%	0.0%	5.7%	1.5%	0.0%	0.0%	3.8%	1.5%
5	3.8%	1.5%	3.8%	0.0%	5.7%	3.1%	0.0%	0.0%	0.0%	0.0%
6	1.9%	3.1%	0.0%	0.0%	3.8%	6.2%	0.0%	0.0%	0.0%	1.5%
7	1.9%	3.1%	0.0%	0.0%	0.0%	3.1%	0.0%	0.0%	0.0%	1.5%
8	3.8%	3.1%	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%
9	1.9%	3.1%	0.0%	0.0%	0.0%	4.6%	0.0%	0.0%	0.0%	0.0%
10+	24.5%	19.6%	5.7%	7.6%	22.5%	15.4%	1.9%	0.0%	1.9%	0.0%
	Control	TXT	Control	TXT	Control	TXT	Control	TXT	Control	TXT
Mean Frequency	14.8	9.0	3.2	1.7	10.2	6.9	0.5	0.1	0.9	0.4
SD	33.7	18.9	10.9	5.9	22.1	14.5	2.9	0.4	4.9	1.3
N	53	65	53	65	53	65	53	65	53	65
Effect size	-0.22		-0.09		-0.21		-0.04		-0.15	
Chi-Sqr	0.03		4.67*		0.09		0.06		0.17	
F-test	1.34		0.91		0.92		1.28		0.71	

* = p-value < 0.05; ** = p-value < 0.01; *** = p-value < 0.001 Effect size: mean Treatment - mean Control/Pooled standard deviation

Table 5: Batterer Reports of Any Domestic Violence Incidents Since Assigned Treatmen

	Prevalence (a)		Rate (b)		Time-to- Failure (c)	
	6 Months		6 Months		To 6 Months	
Type 1	b	Exp(B)	b	Exp(B)	b	Exp(B)
Treatment (yes=1, no=0)	0.069	1.071	-0.489	0.613	-0.068	0.934
Constant	-0.038	0.963	2.692	14.761		
N	118		118		109	
Initial Log likelihood	163.58		318.95		442.85	
Final Log likelihood	163.55		318.93		442.79	
R^2/Cox & Snell pseudo R^2	0.001		0.001		—	
P-value	0.85		0.28		0.81	
Type 2						
Age	-0.014	0.986	-0.015	0.985	-0.002	0.998
Race/Ethnicity (White reference)						
African American	-0.167	0.846	1.322	3.751	0.314	1.369
American Indian	0.767	2.153	0.73	2.075	0.133	1.142
Mixed race	-0.013	0.987	-0.900	0.407	0.050	1.051
Other race	0.604	1.829	1.133	3.105	0.561	1.752
Education - highest completed grade	-0.131	0.877	0.001	1.001	-0.072	0.931
Employed in the 30 days before the arrest (yes=1)	0.254	1.289	0.237	1.267	0.281	1.324
Personal Income	0.000	1.000	-0.001	0.999	0.000	1.000
Relationship Status (Married reference)						
Separated/divorced	-7.283	0.001	-22.042	0.000	-12.41	0.000
Living together	-0.734	0.480	-0.287	0.751	-0.448	0.639
Intimate not living together	-0.985	0.373	-1.397	0.247	-0.517	0.596
Past intimate	-0.718	0.488	-0.85	0.427	-0.139	0.870
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.153	1.165	0.229	1.257	0.208	1.231
Prevalence of physical abuse ever before arrest (self-reported by batterers)	0.004	1.004	0.222	1.249	-0.021	0.979
Prior Arrest for non-domestic violence (yes=1)	-0.006	0.994	-0.129	0.879	-0.020	0.980
Prior Arrest for domestic violence (yes=1)	0.149	1.161	0.479	1.614	0.110	1.116
Treatment (yes=1, no=0)	-0.003	0.997	-0.166	0.847	-0.123	0.884
Constant	1.430	4.179	1.488	4.428		
N	103		103		94	
Initial Log likelihood	142.78		282.57		368.06	
Final Log likelihood	128.66		275.01		354.83	
R^2/Cox & Snell pseudo R2	0.13		0.04			
P-value	0.66		0.15		0.76	

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression

* = p-value < .05; ** = p-value < .01; *** = p-value < .001

Table 6: Batterer Reports Severe DV (Only Physical Abuse) Incidents at Six Months

	Prevalence Severe DV (a) 6 Months			Rate Severe DV (b) 6 Months			Time-to- Failure Severe DV (c) To 6 Months		
Type 1	b	Exp(B)		b	Exp(B)		b	Exp(B)	
Treatment (yes=1, no=0)	-0.232	0.793		-0.358	0.699		-0.526	0.591	
Constant	-1.883	0.152	***	-0.638	0.528				
N	118			118			116		
Initial Log likelihood	85.96			70.01			113.02		
Final Log likelihood	85.79			70.00			112.19		
R^2/Cox & Snell pseudo R^2	0.001			0.001			—		
P-value	0.68			0.66			0.36		
Type 2									
Age	0.122	1.130	*	0.109	1.115		0.136	1.146	**
Race/Ethnicity (White reference)						*			
African American	1.779	5.924		2.72	15.180	*	1.906	6.726	
American Indian	1.347	3.846		1.134	3.108		1.225	3.404	
Mixed race	1.574	4.826		1.094	2.986		1.886	6.593	
Other race	2.121	8.339		1.509	4.522		1.928	6.876	
Education - highest completed grade	-0.243	0.784		-0.517	0.596		-0.287	0.751	
Employed in the 30 days before the arrest (yes=1)	-0.354	0.702		0.053	1.054		0.245	1.278	
Personal Income	0.001	1.001		0.001	1.001		0.001	1.001	
Relationship Status (Married reference)									
Separated/divorced	-8.679	0.000		-18.35	0.000		-15.205	0.000	
Living together	-1.61	0.200		-1.42	0.242		-1.504	0.222	
Intimate not living together	-9.061	0.000		-17.72	0.000		-15.361	0.000	
Past intimate	0.332	1.394		-0.508	0.602		0.197	1.218	
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.090	1.094		0.518	1.679		0.269	1.309	
Prevalence of physical abuse ever before arrest (self-reported by batterers)	0.417	1.517		0.318	1.374		0.407	1.502	
Prior Arrest for non-domestic violence (yes=1)	-0.696	0.499		-1.220	0.295		-0.794	0.452	
Prior Arrest for domestic violence (yes=1)	0.687	1.988		1.220	3.387		0.842	2.321	
Treatment (yes=1, no=0)	0.276	1.318		0.679	1.972		-0.052	0.949	
Constant	-5.070	0.006		-3.220	0.040				
N	103			103			102		
Initial Log likelihood	74.14			62.24			100.69		
Final Log likelihood	51.54			50.57			75.62		
R^2/Cox & Snell pseudo R2	0.20			0.19					
P-value	0.16			0.14			0.33		

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression

* = p-value < .05; ** = p-value < .01; *** = p-value < .001

Table 7: Batterer Reports on Controlling Behavior Incidents at Six Months

	Prevalence Controlling DV (a) 6 Months			Rate Controlling DV (b) 6 Months			Time-to- Failure Controlling DV (c) To 6 Months		
Type 1	b	Exp(B)		b	Exp(B)		b	Exp(B)	
Treatment (yes=1, no=0)	-0.990	0.372	*	-0.646	0.524	***	-1.3313	0.264	**
Constant	-0.838	0.432	***	1.153	3.168	***			
N	118			118			112		
Initial Log likelihood	121.87			670.25			176.10		
Final Log likelihood	117.20			670.00			168.50		
R^2/Cox & Snell pseudo R^2	0.040			0.020			___		
P-value	0.03			0.00			0.01		
Type 2									
Age	0.003	1.003		-0.108	0.898		0.025	1.025	**
Race/Ethnicity (White reference)						*			
African American	0.166	1.180		1.330	3.781	**	0.916	2.500	
American Indian	-20.506	0.000		-15.130	0.000		-13.782	0.000	
Mixed race	-1.018	0.361		-3.304	0.037	**	-13.368	0.000	
Other race	0.339	1.404		1.236	3.442	**	0.346	1.414	
Education - highest completed grade	-0.466	0.628		-0.031	0.969		-0.364	0.695	*
Employed in the 30 days before the arrest (yes=1)	0.135	1.144		0.194	1.214		0.136	1.145	
Personal Income	0.001	1.001		0.000	1.000		0.000	1.000	
Relationship Status (Married reference)		1.000				**			
Separated/divorced	-20.673	0.000		-17.946	0.000		-14.485	0.000	
Living together	-0.437	0.646		-2.046	0.129	***	-0.550	0.577	
Intimate not living together	-1.470	0.230		-2.989	0.050	**	-1.209	0.298	
Past intimate	-1.139	0.320		-1.803	1.106	***	-0.748	0.473	
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.189	1.208		-0.033	0.768		0.358	1.431	
Prevalence of physical abuse ever before arrest (self-reported by batterers)	0.029	1.030		0.101	2.389	*	0.129	1.138	
Prior Arrest for non-domestic violence (yes=1)	-0.576	0.562		-0.264	0.37	**	-0.668	0.513	*
Prior Arrest for domestic violence (yes=1)	0.591	1.807		0.871	111.385	***	0.738	2.091	*
Treatment (yes=1, no=0)	-1.058	0.347	*	-1.003	0.590	***	-1.150	0.317	*
Constant	3.898	49.284		4.713					
N	103			103			98		
Initial Log likelihood	104.20			319.40			144.15		
Final Log likelihood	82.44			308.04			117.35		
R^2/Cox & Snell pseudo R2	0.29			0.50			___		
P-value	0.19			0.00			0.24		

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression
* = p-value < .05; ** = p-value < .01; *** = p-value < .001

Table 8: Frequency of victim reported incidents by assigned treatment: Six month follow-up interview data

	Any DV		Controlling		Psychological Abuse		Threats of Physical Abuse		Physical Abuse	
	Control	TXT	Control	TXT	Control	TXT	Control	TXT	Control	TXT
	%	%	%	%	%	%	%	%	%	%
0	27.5%	32.9%	55.1%	54.1%	33.3%	36.5%	76.8%	74.3%	68.1%	68.5%
1	7.2%	2.8%	5.8%	5.4%	7.2%	0.0%	2.9%	2.7%	4.3%	5.5%
2	4.3%	4.2%	7.2%	1.4%	4.3%	6.8%	4.3%	2.7%	7.2%	2.7%
3	5.8%	4.2%	2.9%	1.4%	4.3%	1.4%	1.4%	2.7%	1.4%	0.0%
4	4.3%	1.4%	1.4%	0.0%	5.8%	2.7%	1.4%	1.4%	1.4%	0.0%
5	0.0%	4.2%	1.4%	0.0%	0.0%	2.7%	1.4%	0.0%	0.0%	0.0%
6	5.8%	0.0%	2.9%	2.7%	1.4%	1.4%	0.0%	1.4%	0.0%	1.4%
7	0.0%	0.0%	1.4%	2.7%	2.9%	0.0%	0.0%	1.4%	1.4%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	0.0%	1.4%	0.0%	1.4%
9	0.0%	5.6%	1.4%	0.0%	0.0%	1.4%	0.0%	0.0%	0.0%	1.4%
10+	45.0%	45.0%	20.0%	32.3%	40.8%	45.7%	11.8%	12.0%	16.2%	19.1%
Mean Frequency	49.7	54.1	11.3	13.8	21.9	24.6	4.4	2.9	11.9	12.7
SD	99.3	93.5	25.9	28.2	32.1	34.5	12.4	7.2	36.4	35.2
N	69	73	69	74	69	74	69	74	69	73
Effect size	-0.22		-0.18		-0.18		-0.20		-0.15	
Chi-Sqr	0.48		0.02		0.16		0.12		0.01	
F-test	0.07		0.29		0.23		0.84		0.02	

* = p-value < 0.05; ** = p-value < 0.01; *** = p-value < 0.001 Effect size: mean Treatment - mean Control/Pooled standard deviation

Table 9: Victim Reports of Any Domestic Violence Incidents Since Assigned Treatment

	Prevalence (a) 6 Months		Rate (b) 6 Months		Time-to- Failure (c) To 6 Months	
Type 1	b	Exp(B)	b	Exp(B)	b	Exp(B)
Treatment (yes=1, no=0)	-0.254	0.776	0.09	1.089	-0.096	0.908
Constant	0.968	2.633	3.91	49.650		
N	142		141		140	
Initial Log likelihood	174.16		580.05		871.59	
Final Log likelihood	173.68		580.05		871.34	
R^2/Cox & Snell pseudo R^2	0.003		0.001		—	
P-value	0.49		0.81		0.64	
Type 2						
Age	-0.056	0.946 *	-0.014	0.986	-0.048	0.953 **
Race/Ethnicity (White reference)						
African American	1.241	3.459	0.146	1.157	0.051	1.052
Asian or Pacific Islander	8.551	5171.894	0.997	2.710	1.08	2.945
American Indian	1.399	4.051	-0.126	0.882	-0.386	0.680
Mixed race	0.196	1.217	-0.366	0.694	-0.548	0.578
Other race	1.278	3.589	-0.497	0.608	0.407	1.502
Education - highest completed grade	-0.158	0.854	-0.259	0.772	0.001	1.001
Employed in the 30 days before the arrest (yes=1)	-0.326	0.722	-0.438	0.645	-0.424	0.654
Personal Income	0.001	1.001 *	0.001	1.001	0.001	1.001 *
Relationship Status (Married reference)						
Separated/divorced	7.692	2190.740	5.728	307.353	2.256	9.545
Living together	-0.358	0.699	0.585	1.795	-0.172	0.842
Intimate not living together	-0.718	0.488	2.321	10.186	-0.288	0.750
Past intimate	-1.224	0.294	0.569	1.766	-0.922	0.398
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.298	1.347	0.236	1.266	0.171	1.186 *
Prevalence of physical abuse ever before arrest (self-reported by batterers)	0.041	1.042	-0.328	0.720	-0.099	0.906
Prior Arrest for non-domestic violence (yes=1)	-0.086	0.918	-0.066	0.936	-0.075	0.928
Prior Arrest for domestic violence (yes=1)	-0.247	0.781	-0.525	0.592	-0.070	0.932
Treatment (yes=1, no=0)	-0.770	0.463	0.257	1.293	-0.300	0.741
Constant	3.606	36.818	6.542	693.669		
N	111		111		110	
Initial Log likelihood	135.10		459.71		654.35	
Final Log likelihood	107.75		456.28		628.52	
R^2/Cox & Snell pseudo R^2	0.22		0.02		—	
P-value	0.07		0.46		0.06	

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression

* = p-value < .05; ** = p-value < .01; *** = p-value < .001

Table 10. Victim Reports Severe DV (Only Physical Abuse) Incidents at Six Months

	Prevalence Severe DV (a) 6 Months			Rate Severe DV (b) 6 Months			Time-to- Failure Severe DV (c) To 6 Months		
Type 1	b	Exp(B)		b	Exp(B)		b	Exp(B)	
Treatment (yes=1, no=0)	-0.017	0.983		0.10	1.101		-0.021	0.979	
Constant	-0.759	0.468	**	0.63	1.868				
N	142			139			141		
Initial Log likelihood	177.36			198.60			421.07		
Final Log likelihood	177.36			198.60			421.06		
R^2/Cox & Snell pseudo R^2	0.001			0.001			——		
P-value	0.96			0.84			0.95		
Type 2									
Age	0.002	1.002		0.034	1.035		0.011	1.011	
Race/Ethnicity (White reference)									
African American	1.163	3.200		0.087	1.091		0.781	2.184	
Asian or Pacific Islander	3.138	23.058		0.374	1.454		1.435	4.200	
American Indian	1.122	3.071		0.202	1.224		0.779	2.179	
Mixed race	-0.027	0.973		-0.063	0.939		-0.17	0.844	
Other race	-0.13	0.878		-0.65	0.522		-0.22	0.803	
Education - highest completed grade	-0.129	0.879		-0.311	0.733		-0.114	0.892	
Employed in the 30 days before the arrest (yes=1)	-1.19	0.304		-0.17	0.844		-0.521	0.594	
Personal Income	0.001	1.001		0.001	1.001		0.001	1.001	
Relationship Status (Married reference)									*
Separated/divorced	9.116	9099.674		7.41	1652.418		4.832	125.461	**
Living together	0.196	1.217		1.33	3.781		-0.081	0.922	
Intimate not living together	1.417	4.125		3.97	52.984		1.262	3.532	
Past intimate	-0.222	0.801		1.66	5.259		-0.099	0.906	
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.258	1.294		0.074	1.077		0.146	1.157	
Prevalence of physical abuse ever before arrest (self-reported by batterers)	-0.076	0.927		-0.389	0.678		-0.053	0.948	
Prior Arrest for non-domestic violence (yes=1)	-0.142	0.868		-0.159	0.853		-0.118	0.889	
Prior Arrest for domestic violence (yes=1)	-0.283	0.754		-0.615	0.541		-0.162	0.850	
Treatment (yes=1, no=0)	-0.315	0.730		0.169	1.184		-0.104	0.901	
Constant	0.637	1.891		2.064	7.877				
N	111			109			111		
Initial Log likelihood	138.37			153.45			318.24		
Final Log likelihood	117.32			148.20			299.36		
R^2/Cox & Snell pseudo R^2	0.17			0.05					
P-value	0.28			0.70			0.01		

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression

* = p-value < .05; ** = p-value < .01; *** = p-value < .001

Table 11 Victim Reports on Controlling Behavior Incidents Committed by Batterer at Six Months

	Prevalence Controlling DV (a) 6 Months			Rate Controlling DV (b) 6 Months			Time-to- Failure Controlling DV (c) To 6 Months	
Type 1	b	Exp(B)		b	Exp(B)		b	Exp(B)
Treatment (yes=1, no=0)	0.041	1.042		0.028	1.028		0.053	1.054
Constant	-0.204	0.815		2.429	11.348			
N	143			143			139	
Initial Log likelihood	197.06			2762.04			580.59	
Final Log likelihood	197.04			2762.04			580.49	
R^2/Cox & Snell pseudo R^2	0.000			0.006			___	
P-value	0.90			0.00			0.83	
Type 2								
Age	-0.080	0.923	*	-0.016	0.984	***	-0.030	0.970
Race/Ethnicity (White reference)								
African American	1.587	4.889		0.579	1.784		0.819	2.268
Asian	2.314	10.115	*	0.633			0.225	
American Indian	0.441	1.554		-1.145	1.883		-0.247	0.781
Mixed race	-0.831	0.436		-0.471	0.318		-0.688406	0.502
Other race	0.022	1.022		-1.442	0.236		-0.647309	0.523
Education - highest completed grade	-0.303	0.739		-0.370	0.691	***	-0.209175	0.811
Employed in the 30 days before the arrest (yes=1)	-2.056	0.128	**	-0.522	0.593	***	-0.893267	0.409
Personal Income	0.000	1.000		0.000	1.000		0.001	1.001
Relationship Status (Married reference)		1.000						
Separated/divorced	23.921	24476612566.7		5.101	164.185		4.3180872	75.045
Living together	-0.009	0.991		0.357	1.429		-0.028503	0.972
Intimate not living together	1.141	3.130		1.269	3.557		1.1067793	3.025
Past intimate	-0.617	0.540		-0.542	0.582		-0.088719	0.915
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.148	1.160		0.066	1.068	***	0.0633646	1.065
Prevalence of physical abuse ever before arrest (self-reported by batterers)	-0.316	0.729		-0.224	0.799	***	-0.272528	0.761
Prior Arrest for non-domestic violence (yes=1)	-0.067	0.935		-0.133	0.88	***	-0.083844	0.920
Prior Arrest for domestic violence (yes=1)	-0.373	0.689		-0.321	0.725	***	-0.182848	0.833
Treatment (yes=1, no=0)	-0.437	0.646		0.048	1.049		-0.156962	0.855
Constant	6.954	1047.326	**	7.030	1130.025	***		
N	112			112			109	
Initial Log likelihood	154.37			5562.75			435.14	
Final Log likelihood	116.85			1541.78			404.62	
R^2/Cox & Snell pseudo R2	0.29			0.34			___	
P-value	0.004			0.001			0.007	

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression

* = p-value < .05; ** = p-value < .01; *** = p-value < .001

Table 12: Frequency of arrests for domestic violence by Assigned Treatment

	Officially Recorded Incidents Six Months		Officially Recorded Incidents Twelve Months	
	Control	TXT	Control	TXT
	%	%	%	%
0	34.3%	34.1%	31.4%	30.4%
1	48.1%	48.1%	45.6%	45.5%
2	13.8%	11.7%	14.1%	13.7%
3	3.2%	5.4%	6.4%	7.4%
4	0.4%	0.7%	1.0%	2.7%
5	0.4%	0.0%	0.7%	0.3%
6	0.0%	0.0%	0.4%	0.0%
7	0.0%	0.0%	0.4%	0.0%
8	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%
10+	0.0%	0.0%	0.0%	0.0%
Mean Frequency	0.88	0.90	1.05	1.07
SD	0.82	0.85	1.07	1.01
N	312	317	312	317
Effect size	0.02		0.02	
Chi-Sqr	0.002		0.07	
F-test	0.08		0.06	

* = p-value < 0.05; ** = p-value < 0.01; *** = p-value < 0.001

Effect size: mean Treatment - mean Control/Pooled standard deviation

Table 13: Officially Recorded Arrests for Any Domestic Violence Since Assigned Treatment

	Prevalence (a)				Rate (b)				Time-to-Failure (c)	
	6 Months		12 Months		6 Months		12 Months		To 12 Months	
Type 1	b	Exp(B)	b	Exp(B)	b	Exp(B)	b	Exp(B)	b	Exp(B)
Treatment (yes=1, no=0)	0.007	1.007	0.047	1.048	0.022	1.022	0.019	1.019	0.031	1.031
Constant	0.651	1.917 ***	0.779	2.179	-0.124	0.883	0.052	1.053		
N	582		582		582		582		582	
Initial Log likelihood	747.64		719.96		688.95		773.04		4810.32	
Final Log likelihood	747.64		719.89		688.94		773.04		4810.22	
R^2/Cox & Snell pseudo R^2	0.0001		0.001		0.001		0.001			
P-value	0.97		0.79		0.80		0.81		0.76	
Type 2										
Age	0.01	1.010	0.007	1.007	0.003	0.998	0.003	1.003	0.004	1.004
Race/Ethnicity (White reference)										
African American	-0.372	0.689	-0.404	0.668	-0.23	0.795	-0.111	0.895	-0.307	0.736
Asian or Pacific Islander	-0.76	0.468	-0.922	0.398	-0.634	0.530	-0.775	0.461	-0.633	0.531
American Indian	-0.012	0.988	0.161	1.175	0.178	1.195	0.267	1.306	-0.092	0.912
Mixed race	0.12	1.127	0.183	1.201	0.035	1.036	0.11	1.116	0.056	1.058
Other race	-0.268	0.765	-0.191	0.826	-0.072	0.9305	-0.065	0.9371	-0.177	0.838
Education - highest completed grade	0.046	1.047	0.050	1.051	0.027	1.027	0.025	1.025	0.016	1.016
Employed in the 30 days before the arrest (yes=1)	0.328	1.388	0.304	1.355	0.158	1.171	0.124	1.132	0.186	1.204
Personal Income	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Relationship Status (Married reference)										
Separated/divorced	-0.086	0.918	-0.135	0.874	0.014	1.014	0.051	1.052	0.084	1.088
Living together	-0.293	0.746	-0.31	0.733	-0.082	0.921	-0.039	0.962	-0.169	0.845
Intimate not living together	-0.157	0.855	-0.043	0.958	-0.141	0.868	-0.069	0.933	-0.038	0.963
Past intimate	-0.349	0.705	-0.122	0.885	0.038	1.039	0.056	1.058	-0.063	0.939
Prevalence of psychological abuse ever before arrest (self-reported by batterers)	0.078	1.081	0.05	1.051	0.018	1.018	0.014	1.014	0.008	1.008
Prevalence of physical abuse ever before arrest (self-reported by batterers)	0.007	1.007	0.006	1.006	0.026	1.026	0.043	1.044	0.01	1.010
Prior Arrest for non-domestic violence (yes=1)	0.077	1.080	0.067	1.069	0.049	1.050 *	0.054	1.055 *	0.031	1.031
Prior Arrest for domestic violence (yes=1)	-0.104	0.901	-0.015	0.985	-0.028	0.972	-0.014	0.986	-0.037	0.964
Treatment (yes=1, no=0)	-0.105	0.900	-0.063	0.939	0.048	1.049	0.066	1.068	0.001	1.001
Constant	-0.125	0.882	0.052	1.053	-0.66	0.517	-0.51	0.599		
N	461		461		461		461		461	
Initial Log likelihood	583.10		566.07		526.06		595.35		3691.98	
Final Log likelihood	566.46		552.39		525.95		595.15		3677.89	
R^2/Cox & Snell pseudo R2	0.0350		0.03		0.02		0.02			
P-value	0.55		0.75		0.42		0.11		0.75	

a=Logistic Regression; b=Negative Binomial Regression; c=Cox Regression

* = p-value < .05; ** = p-value < .01; *** = p-value < .001