Document Title:	Does Batterer Treatment Reduce Violence? A Randomized Experiment in Brooklyn – Executive Summary Included
Author(s):	Robert C. Davis ; Bruce G. Taylor ; Christopher D. Maxwell
Document No.:	180772
Date Received:	February 8, 2000
Award Number:	94-IJ-CX-0047

180772

Does Batterer Treatment Reduce Violence?

A Randomized Experiment in Brooklyn

Robert C. Davis

Bruce G. Taylor

Christopher D. Maxwell

Victim Services Research
346 Broadway, Suite 206
NY, NY 10013

January 3, 2000

ABSTRACT

During the past two decades, pro-arrest laws have resulted in an increasing number of prosecutions of men who assault spouses or girlfriends. Researchers and practitioners have documented the difficulty of altering the behavior of convicted spouse abusers. As the courts have searched for effective sanctions for spouse abusers, they have increasingly come to rely on group treatment programs as the sentence of choice for the widening pool of men convicted of spousal assault.

The greater reliance on batterer treatment programs makes it important that we can document that such programs effectively reduce the propensity of offenders to commit new violence. There is no shortage of evaluations of batterer treatment programs: Some three dozen have appeared in the literature since the 1980s. Most of these studies have methodological deficiencies, which make it difficult to interpret their findings. But evaluation studies have become more sophisticated as time has passed.

The present study represents one of the first attempts to conduct a test of batterer treatment using a true experimental design. The design randomly assigned 376 court-mandated batterers to batterer treatment or to a treatment irrelevant to the battering problem (community service). All men assigned to batterer treatment were mandated to 39 hours of class time. But some were assigned to complete the treatment in 26 weeks and others in eight weeks. Men assigned to the control condition were sentenced to forty hours of community service. For all cases in the study, interviews were attempted with victims and batterers at 6 months and 12 months after the sentence date. In addition, records of criminal justice agencies were checked to determine if new crime reports or arrests had occurred involving the same defendant and victim.

The results showed that treatment completion rates were higher for the eight-week group than for the 26-week group. However, only defendants assigned to the 26-week group showed significantly lower recidivism at 6 and 12 months post-sentencing compared to defendants assigned to the control condition. The groups did not differ significantly at either 6 or 12 months in terms of new incidents reported by victims to research interviewers. We interpret the results to indicate that batterer intervention has a significant effect on suppressing violent behavior while batterers are under court control, but may not produce

INTRODUCTION

Over the past two decades, the law enforcement response to domestic violence has become increasingly tough. 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 domestic violence 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 group treatment programs as the sanction of choice.

There are compelling reasons why group treatment programs for batterers have become a popular mode of court sanction. Even in serious battering cases, many victims choose to stay with abusive partners. Such victims are interested in sanctions which offer them safety from violence, not retribution or punishment that will jeopardize their partner's ability to earn a living. Alternative sanctions commonly used in other crimes have little face validity in abuse cases: There is little reason to believe that fines, community service or probation without special conditions will stop batterers from abusing their spouses.

There is no shortage of evaluations of batterer treatment programs. But the vast majority has serious methodological flaws which make it impossible to distinguish between treatment effects, temporal effects, and selection effects. Generally, the evaluation literature shows an evolution toward more rigorous science since the first batterer treatment studies appeared in the literature in the early 1980s. The study we describe represents one of the first attempts to conduct a test of batterer treatment using a true experimental design which randomly assigns court-mandated batterers to batterer treatment or to a control condition.

The Nature of Batterer Treatment

The first group programs for batterers were begun during the late 1970s. Feminists, victim advocates, and others realized that providing services to victims of abuse and then returning them to the same home environment did little to solve abuse problems (Healey, Smith, and O'Sullivan, 1997). Group treatment was believed to be more appropriate than individual counseling or marital therapy because it expanded the social networks of batterers to include peers who are supportive of being nonabusive (Crowell and Burgess, 1996). Groups also proved to be less expensive than one-on-one counseling sessions. The earliest batterer groups were educational groups which sought to promote an anti-sexist message (Gondolf, 1995). With the passage of time, they gradually incorporated cognitive/behavioral therapeutic techniques and skill-building exercises.

As states introduced pro-arrest statutes during the 1980s the number of batterers arrested and convicted increased, and group treatment became the treatment of choice for the courts. Court-mandated batterer treatment significantly increased and diversified the number of batterer programs nationally (Feazell, Mayers, & Deschner, 1984). A recent estimate places the proportion of court mandates in treatment programs at 80% (Healey, et. al. 1997).

Batterer treatment may be required by criminal courts as part of a pre-trial diversion program, may be ordered by judges as part of a sentence, or may be imposed by probation agencies empowered to set special conditions of probation (Hamberger & Hastings, 1993). In at least one major urban jurisdiction, the district attorney sometimes agrees not to file charges at all if a brief treatment program is completed (Davis and Smith, 1997). In some states (see Ganley, 1987), civil courts as well as criminal may mandate a batterer to treatment (e.g., as a condition related to child visitation).

Many batterer programs are run by probation departments, while others are run by mental health practitioners, family service organizations, or victim service programs. Intake practices vary, with some programs accepting all court referrals and others exercising discretion in excluding persons with prior convictions or substance abuse problems. Supervision of batterers in treatment can most often falls to probation officers, but is sometimes undertaken by others - and increasingly by judges. Historically, supervision has been lax, drop out rates high, and sanctions unevenly applied. Recently, however, supervision has become stricter and sanctions for failure to attend sessions more common.

Program Typologies

Different perspectives on wife battering place the cause within individuals (personality or psychological abnormalities of batterers), within family dynamics (dysfunctional communication), or within the community (societal attitudes supporting violence). There are a wide variety of batterer treatment programs which address several of these three different levels of causation. Adams (1988) and Hamberger and Hastings (1993) differentiate batterer treatment groups according to five philosophical orientations. The *feminist* framework is a political approach which proposes that male-to-female violence is rooted in a patriarchal society which provides power to men and oppresses women (Hamberger and Hastings, 1993). Domestic violence is seen as a means of establishing and maintaining male dominance, and is viewed as a by-product of male and female sex roles. Subordinate economic roles have made women dependent on men and unable to leave their abusive situation. Feminist-based treatment programs rely primarily on "re-educating" batterers about the roles of men and women and about appropriate behavior in intimate relationships.

The *cognitive-behavioral* model, based on social learning theory, views domestic violence as behavior learned by batterers through direct observation of role models, indirect observation (e.g., through the media), and direct "trial and error" learning experiences (Hamberger and Hastings, 1993: 199). Violence is seen as functional for the perpetrator (e.g., tension release, avoidance of unpleasant tasks, and enforced victim compliance). Batterer groups based on this model teach batterers conflict avoidance techniques, assertiveness skills, relaxation skills, and cognitive strategies for reevaluating and neutralizing anger-producing thoughts.

The *ventilation* model views partner violence as symptomatic of suppressed anger that needs to be expressed through some other means This model is rooted in family dynamics and views both partners as responsible for the violence. Batterers, and often their partners as well are assigned to groups which work on developing better communication within the dyadic relationship.

The *insight-oriented* model views domestic violence as a symptom of underlying problems from the batterer's past (e.g., residual fear or anger from past abuse from parents) that unconsciously motivates current violent behavior (Hamberger and Hastings, 1993: 197). Treatment involves examining inner-life experiences, past experiences, and current interactions with others.

The *systems* model is based on the idea that domestic violence is spawned by competition for control in dyadic relationships in which each partner attempts to dominate and control the other (Hamberger and Hastings, 1993). The early stages of this process begin with verbal and emotional abuse, but as both partners strive to win, one of the partners may resort to violence. Both parties participate in groups together. The group works on helping each partner identify their role in the violence, and improving communication skills (Adams, 1988).

In practice, modern batterer groups tend to mix different theoretical approaches to treatment (Healey, et. al. 1997), although most batterer programs are based upon the feminist model developed by the Domestic Abuse Intervention Project in Duluth, Minnesota. The Duluth model assumes that physical violence is part of a spectrum of male efforts to control women. But batterer programs also commonly deal with the need for anger control, stress management, and better communication skills.

Not only treatment approach, but treatment length varies from program to program. The duration or number of sessions may vary from as little as one day to 32 weeks (Feazel et al., 1984). Some in the field even have advocated long-term treatment from 1 to 5 years (Ewing, Lindsey, & Pomerantz, 1984). However, there also is substantial pressure to keep batterer treatment short in duration resulting from pressure from insurance companies' imposition of time limits for batterers seeking reimbursement (Edelson and Syers, 1990).

Current trends in treatment programs seem to be going in conflicting directions. Increasingly, states are developing guidelines to codify standards for treatment content and length among batterer treatment programs (Gondolf, 1995). But, on the other hand, there is increasing sentiment that a "one-size fits all" approach to batterer treatment fails to recognize the diversity of batterers that enter treatment (Healey, et. al. 1997). There is a trend for treatment programs to tailor interventions to different batterer types defined by personality, violence history, or substance abuse. Other programs have been specially designed to accommodate sociocultural differences among batterers such as poverty, ethnicity, or sexual orientation.

The Evaluation Literature

Over the last two decades there have been many empirical studies on batterer treatment programs. There are at least six published reviews of over 35 published single-site evaluations (e.g., Eisikovits & Edleson, 1989; Gondolf, 1991,1995; Rosenfield, 1992; Saunders, 1996a; Tolman & Bennett, 1990) and eight research reviews (e.g., Davis and Taylor, in press; Hamberger & Hastings, 1993; Crowell & Burgess, 1996; Dobash, Dobash, Cavanagh & Lewis, 1995; Dutton, 1988, 1995; Rosenbaum & O'Leary, 1986; Saunders & Azar, 1989; Tolman & Edleson, 1995). Since these literature reviews a number of new studies have been conducted and published.

However, the volume of the literature is deceptive. In fact, there have been only a handful of investigations that can make any legitimate claims about differences between treated batterers and untreated batterers. The batterer treatment literature has gone through three generations of studies. Most recent have been investigations which have randomly assigned batterers to treatment conditions. These are the strongest designs. Quasi-experiments of varying quality appeared somewhat earlier in the literature. The oldest, and by far the largest, portion of the empirical literature consists of studies which examine only batterers assigned to treatment programs. Included in this set of studies are: (a) studies which assess violence or other individual outcomes only after batterer treatment, (b) studies which measure violence before and after treatment, and © studies which compare violence of batterers who complete treatment with batterers assigned to treatment, but do not attend. Although the methodologies of early studies do not tend to be strong, they are important because they laid the foundation upon which stronger designs could be developed.

Methodological Issues in the Literature

In order to intelligently evaluate treatment outcome studies, it is important to have in mind some of the methodological shortcomings common in this literature. This section outlines some of the major problems which are common to many studies. In the reviews which follow this section, we will draw upon this understanding to evaluate particular investigations and groups of studies.

First, there has been a lack of consensus on how to measure program effects. Studies have measured program effects on violence using official data on arrests and complaints, victim surveys, and batterer surveys. Rosenfeld's (1992) review makes the point in detail that official reports of violence and batterer surveys seriously underestimate actual violence committed in relationships. Moreover, some studies (e.g. Mauiro, Cahn, Vitaliano, and Zegree 1987) have not included any indicators of violence in their outcome measures. (Such studies are not included in our review.) Follow-up intervals have varied greatly, from several months to several years.

Studies differ widely in their statistical sophistication. While most have reported inferential statistics examining differences between means, a few have merely presented percentage differences. Some studies which did use inferential statistics were conducted without sufficient statistical power to detect differences between treated and untreated participants. Some of the best quasi-experiments have incorporated multivariate analyses which attempt to control for the effects of extraneous variables when isolating treatment effects.

Studies have varied in terms of the populations they are investigating. Obviously, the samples in these studies are not going to representative of all batterers in the United States, or even all batterers mandated to batterer treatment in the United States. Most researchers would probably be satisfied with demonstrating that batterer programs are effective for some well-defined group of batterers in one court system, in one city. Clearly, obvious sample selection biases should be avoided.

One such sample selection bias is that most of the batterer programs that have been evaluated exclude difficult batterers (e.g., recidivist batterers or those who have substance abuse problems) from their programs. Elimination of potentially difficult subjects may overestimate the successfulness of treatment programs, were these programs forced to accept these more difficult cases (Rosenfeld, 1992). Therefore, the results of many of these studies may apply only to a limited spectrum of batterers.

The problem of generalizeability of results also crops up in another way. Many treatment studies which have relied on batterer or victim surveys to assess violence have had poor interview response rates, some as low as 30%. Low response rates are a problem because the cases in which follow-up data are available may be different than those cases 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 compared to batterers who did not do a follow-up survey. 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 complete it.

Finally, batterer treatment programs have serious problems with attrition: Many evaluations report that fewer than half of batterers assigned to treatment, in fact, completed the program. Low treatment completion rates present researchers conducting experiments or quasi-experiments with a dilemma. If they compare only batterers who complete treatment with batterers not assigned to treatment, they are subject to criticisms of "creaming". That is, they are comparing the best of the treatment group (the most highly motivated batterers) with untreated batterers, thereby stacking the deck in favor of finding program effects. On the other hand, if all batterers assigned to treatment are included in the comparison, yet most failed to complete treatment, they are subject to the criticism that their study is biased against finding program effects. In other words, program effects would have to be very large, indeed, to show up after being diluted by inclusion of drop-outs who were not exposed to the treatment (or exposed to a lesser treatment "dosage").

Studies Without a Comparison Group

Non-experimental one group post-test only designs

At least 15 published studies have used designs which generate a single measure of treatment effectiveness: violence following completion of treatment (see Table 1). Ten measured recidivism based only upon batterer self-reports. Only four of the fifteen studies had substantial sample sizes (which we have arbitrarily defined as greater than 100) or lengthy follow-up periods (which we have defined as one year or greater).

Recidivism rates in this group of studies vary widely, from 7% to 47% (mean 26%). Interpretation of results is difficult at best without a comparison group or pre-test information with which to compare outcome measures.

Non-experimental one group pre-test and post-test designs

At least seven published studies compared violence among treated batterers after program participation to violence levels prior to participation (see Table 2). Three of the seven studies included both victim and batterer self-reports, but just two had follow-up periods of at least a year and none of the studies examined police records. Two of the seven studies had sample sizes greater than 100. Of the six studies that reported treatment attrition rates, four of the studies had attrition rates of 25% or

Table 1: Batterer Treatment Evaluations Using a Post-Test Only Design

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivisim	Attrition
Purdy & Nickle (1981)	170	Batterer	6 months	41%	Unknown
Deschner (1984)	12	Batterer	8 months	15%	50%
Feazel, Mayers, and Deschner (1984)	90	Batterer	1 Year	25%	Unknown
Edleson, Miller, Stone, and Chapman (1985)	9	Batterer	7 to 21 weeks	22%	0%
Neidig, Friedman, and Collins (1985)	Unknown	Batterer	4 months	13%	Unknown
Harris (1986)	40	Batterer	2 months to 3 years	27%	Unknown
DeMaris and Jackson (1987)	53	Batterer	20 months	35%	83%
Leong, Coates, and Hoskins (1987)	67	Victim, Police	3 months	19% (Victim) 15% (Police)	76%
Shupe, Stacey & Hazlewood (1987)	148	Victim, Batterer	3 months to a few years	30% (Victim) 18% (Batterer)	31%
Tolman, Beeman, and Mendoza (1987	48	Victim	6 months	47%	68%
Edleson and Grusznski (1988) (Study 2)	86	Victim	9 months	33%	0%
Beninati (1989)	16	Batterer	Unknown	19%	25%
Hamberger and Hastings (1990)	106	Batterer	1 year	30%	16%
Johnson and Kanzler (1990)	687	Batterer	5 months	7%	30%
Tolman and Bhosley (1991)	99	Victim	1 year	42%	50%

Table 2: One Group Pre and Post-Test Design

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivisim	Attrition
Dutton (1986) Part 1	50	Batterer & Victim	6 months to 3 years	Pre-Test 13.4 All DV acts (Batterer reports) / Post-Test 4.6 All DV acts (Batterer reports) Pre-Test 21.3 All DV acts (Victim reports)/ Post-Test 6.1 All DV acts(Victim reports) (For all differences, P<.05)	10%
Rosenbaum (1986)	11	Batterer	4 & 6 Months	100% (Pre-treatment) 9% (4 months) 27% (6 Months) (P< .05)	18%
Waldo (1986)	23	Batterer	6 Months	Pre-Test 5.1 DV acts / Post-Test 0.29 DV acts (P < .05)	Unknown
Shepard (1987)	92	Batterer	14 months	Pre-Test 39% / Post Test 30% (Statistical significance not reported)	25%
Hamberger and Hastings (1988) Part 1	35	Batterer, Victim (Combined measure)	1 year	Pre-Test 20.9 DV acts / Post-Test 5.3 DV acts (P < .001)	0%
Meredith & Burns (1990)	125	Batterer, Victim	3 months	Physical, verbal & emotional abuse all reduced at post-test (% not reported)	53%

less.

All seven studies reported lower recidivism rates following treatment (but results of one study were not statistically significant; two studies did not report probability statistics). However, with this type of design, reductions in recidivism cannot be attributed necessarily to the effects of treatment. This is true because studies have repeatedly 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 repeat domestic violence within the next six months after the police intervene (see, for example, Davis and Taylor, 1997; Sherman, 1992; Fagan, Friedman, Wexler, and Lewis 1984). The post-treatment violence rates displayed in Table 2 also average about one-third -- in other words not different than one might expect even if the batterers had not undergone treatment.

Comparing treatment drop-outs versus completers Six studies compared outcomes between batterers who completed treatment and batterers assigned to a treatment program, but who failed to complete treatment (see Table 3). Four of the six studies had sample sizes under 100. Only two of the six studies had follow-up periods of at least one year, and just one included more than a single measure of recidivism.

The most serious flaw in these six studies is that the treated and untreated (dropout) groups are almost certainly not comparable in complex ways prior to treatment. As pointed out by Palmer, Brown, and Barrera (1992), attendance is a confounding factor

Table 3: Quasi-Experiment (Dropouts Versus Completers)

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivism
Halpern (1984)	84	Victim	3 months	18% dropouts / 15% completers (N.S.)
Hawkins & Beauvais (1985)	106	Police	6 months	18% Dropouts / 18% completers (N.S.)
Douglas & Perrin (1987)	40	Police	6 months	29% Dropouts / 15% Completers (No Statistics Reported)
Edleson and Grusznski (1988, Study1)	86	Victim	About 5 to 9 months	46% Dropouts / 32% completeres (P < .03)
Edleson and Grusznski (1988, Study 3)	159	Victim	1 year	48% Dropouts / 41% completers (N.S.)
Hamberger and Hastings (1988) Part 2	71	Batterer, Victim, Police (Combined measure)	1 year	47% dropouts / 28% completers (P <.06)

because better attendance is likely an indication of higher motivation to change, even before treatment. Therefore, differential recidivism between program completers and drop-outs could be due to motivational differences in the two groups that existed prior to treatment. Surprisingly, however, only one of the six studies reported significantly lower recidivism rates for the completers (four of the other five studies were in the predicted direction but either had results that were not statistically significant or did not include inferential statistics).

The best use of this group of studies is to describe the characteristics of people that drop-out of treatment -- information potentially useful to program developers to improve batterer groups. Results have indicated that those who do not complete treatment are more likely to be victims of child abuse (Grusznki & Carrillo, 1988), unemployed (Hamberger & Hastings, 1988;), uneducated (Grusznki & Carrillo, 1988), young (Hamberger & Hastings, 1993), psychologically disturbed (Hamberger & Hastings, 1989; Grusznki & Carrillo, 1988), and substance abusers (Hamberger & Hastings, 1990).

Quasi-Experimental Non-Equivalent Matched Groups

We found four studies in which batterers mandated to treatment by the courts were compared to batterers who received other interventions. This group of studies is the first we have examined which addressed in a rigorous fashion the issue of whether treatment works. There is a notable difference in design details between these four quasi-experiments and the other studies reviewed thus far. All four of the studies had sample sizes greater than 100 (see Table 4). None of the studies relied solely on batterer self-reports. All four had follow-up periods of at least one year.

The first quasi-experiment was reported by Dutton (1986). His sample consisted of 100 convicted batterers on probation. He compared 50 batterers who were treated within a cognitive-behavioral group model to 50 batterers who were not designated to receive treatment. The treatment group had a 4% recidivism rate compared to 40% for the control group based upon police reports. However, although Dutton reports that groups did not differ on several demographic measures, pre-treatment comparability of the groups is highly suspect: The control group was composed of batterers whom probation officers did not select for treatment, some of whom were explicitly rejected by therapists as unsuitable for treatment. The treatment group consisted of only batterers who *completed* the treatment program. Dutton does not report what proportion of all batterers assigned to treatment dropped out but, based on other work, we have to assume that it was a large proportion.

Chen et al. (1989) conducted a quasi-experiment involving 120 batterers assigned to treatment by the courts and 101 comparison batterers drawn from court calendars who were not mandated to go to treatment. (No details are given on how the controls were selected or what the outcomes were of their court cases, although the authors state that the samples proved to be well-matched

Table 4: Quasi-Experiment (Matched Control Group)

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivisim	Attrition
Dutton (1986) Part 2	100	Police	6 months to 3 years	40% No treatment / 4% Treatment (P < .001)	0%
Chen, Bersani, Myers, and Denton (1989)	221	Police	Average of 14 months	10% (0.53 DV acts) No Treatment / 5% (0.35 DV acts) Treatment (P < .05) Perps Attended >75% TX less recidivisim than controls(P<.05)	Unknown
Harrell (1991)	348	Batterer/ Victim (Combined measure), Police	6 months for batterer & victims, 15 and 29 Months for police	15% severe violence No Treatment / 20% Treatment (P=N.S.), 12% physical aggression No TX / 43% Treatment (P<.01) 7% New DV Charges No Treatment / 19% Treatment (P < .05)	24%
Dobash et al (1996)	313	victim & court reports	3 & 12 months	7% treated, 10% untreated (court reports 12 months) 30% treated, 62% Untreated (victim 3 months) 33% treated, 75% untreated (victim 12 months) No probability statistics provided	Unknown

demographically.). Sixty-three percent of the men assigned to treatment completed at least 75% of the required sessions. Chen et al. also used a sophisticated data analytic technique (selection bias modeling) to deal with the potential non-equivalence of groups prior to treatment inherent in non-randomized experiments. They found that, after an average of 14 months, 5% of batterers assigned to treatment had been rearrested compared to 10% of controls. The main effect of the treatment variable was not statistically significant, although the authors noted that batterers who completed at least 75% of the requisite sessions had significantly lower rates of recidivism than controls.

Harrell (1991) studied 227 batterers, 115 of whom were ordered to treatment by judges. (She does not specify what court outcomes of the untreated group were.) Her attempt to obtain equivalency between those treated batterers controls hinged on a quirk in the court she studied. She noted that treatment program referrals came almost exclusively from a small group of judges; other judges seldom mandated treatment for batterers. Therefore, she drew her comparison group from the caseloads of judges who seldomly referred to the treatment program. However, her plan did not work as she had intended. Harrell found three important and statistically significant differences between treated batterers and controls. (The former were more likely to be married to their partners and employed, and less likely to have a criminal record). While she controlled for these variables in her analysis of recidivism effects, it is quite possible that there were additional, unmeasured differences between the groups.

Harrell's analysis included only batterers in the treatment group who actually completed treatment. Comparisons of recidivism were based on a combined measure of the victim and perpetrator reports of violence six months after case disposition. In addition, police records were reviewed 15-29 months after case disposition. Surprisingly, a significantly larger percentage of those in the treatment group committed new violence than those in the control group for two of three measures that she reports. (The third measure is in the same direction, but not statistically significant.). For example, 7% of the control group and 19% of the treatment group were charged with new domestic crimes. While Harrell's study may be limited in its ability to distinguish between selection effects and treatment effects, it certainly adds controversy to the debate about the efficacy of treatment programs.

Recently, Dobash, Emerson-Dobash, Cavanagh and Lewis (1996) reported on a quasi-experiment evaluating a treatment program in Great Britain. Dobash et al. 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. (These differences are reminiscent of pre-treatment differences in Harrell's study.)

It is not specified whether Dobash, et. al. included in their analyses all batterers assigned to treatment, or only those who completed treatment. According to court reports at 12 months follow-up, 7% of the treatment group recidivated compared to 10% of the control group: No statistical tests were reported to indicate whether the difference was significant. 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, although no specific information is provided.) However, the success rate for interviews was low: Dobash et al. interviewed only 43% of the victims at the first follow-up interview, 34% at the second interview, and 25% at the third interview.

Randomized Experiments

As pointed out by Palmer et al. (1992), quasi-experiments on batterer treatment cannot be relied upon to produce unbiased estimates of the effects of treatment. This is true because we cannot know whether batterers assigned to treatment and controls are equivalent prior to application of the treatment. In some quasi-experiments (such as the Dutton, 1986 or Harrell, 1991 studies), we know for certain that selection bias favored finding treatment effects (because the control group was comprised of batterers more prone to recidivate than those in the treated group).

It can be argued that initial differences between groups can be controlled statistically, but this is only true if all relevant initial differences are known to researchers. For example, a researcher may discover pre-treatment differences in employment, marital status, and criminal history between those assigned to batterer treatment and controls, and these differences may be statistically controlled in analyses. However, groups may well have differed on less tangible and more fundamental factors such as emotional maturity as well. If such factors are not controlled (because they are not known) and they are correlated with outcome measures, then the results of the study are uninterpretable. The safest way to ensure that estimates of sample means are unbiased is through random assignment of batterers to treatments.

Palmer et. al. conducted the first experiment with random assignment to a true no treatment control group. The number of subjects in the experiment was far smaller than one would expect to need to detect treatment effects: Fifty-nine probationers were assigned using a "block random" procedure to either a ten-session psychoeducational group (combining group discussion with information) or a no treatment control group: Participants were assigned to treatment if a new group was to commence within three weeks; otherwise they became part of the control group. In only two cases was a defendant assigned to the control condition reassigned by court officials to the treatment condition. Attrition was kept within a respectable range: 70% of the men assigned to treatment attended at least seven of the required 10 sessions.

It is significant that this is one of the only studies to compare all batterers *assigned* to treatment (not just those who completed treatment) with controls. Palmer and her colleagues examined police reports six months post-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 (10% compared to 31%). Even with the small N, this difference was statistically significant. While Palmer et. al. attempted to generate additional violence measures from surveys of interviews and batterers, low response rates combined with a small N preçluded any analysis of recidivism based upon interview data.

Two additional randomized experiments are in progress. Dunford (1997) is in the final stages of comparing treatment outcomes for 861 legally married Navy couples in which physical abuse had come to the attention of Navy authorities. These cases were randomly assigned to one of four treatments, including (a) 26-week batterer treatment (based on a cognitive/.behavioral model), (b) 26 weeks of couples counseling, © rigorous monitoring (including monthly calls to victims and semi-annual police record checks), and (d) establishing a safety plan for victims. The safety planning was intended by the investigators as a no-treatment control against which to compare the effects of the other three treatments. (Safety planning was given to victims in each of the other three conditions as well.) This would seem to be a fairly good no-treatment condition, in so far as the men in this condition received no intervention. Victims and batterers are being interviewed every six months over a period of two years. Feder (1996) has assigned batterers placed on probation to either a 26-

week educational batterer program based on the Duluth model or a control group not mandated to treatment. Multiple measures of recidivism will be assessed (victim, batterer, police records, probation records) for six months and one year.

Purposes of the Present Study

We sought to add to the incipient literature on randomized studies of batterer treatment. Although any form of design can be criticized, we concur with Fagan (1996) that randomized experiments entail less serious problems than other designs. A properly executed randomization process is the only way to ensure that treatment effects are not confounded with pre-existing subject characteristics. Our study adds to the literature on randomized experiments in several important ways.

Unlike the sites of the Palmer and Feder experiments, batterers in the site of our study were mandated to treatment by judicial order (in the sites of the other two studies, orders to treatment were made by probation departments). This difference has implications for the kinds of batterers studied. The Palmer and Feder studies had a wide sampling frame, including all or most batterers sentenced to probation, regardless of the batterers' willingness or unwillingness to enter into treatment. In our study, batterers were only eligible for inclusion if all parties to the case (prosecution, defense, and judge) agreed that treatment was appropriate. Such agreement was forthcoming in a small percentage of cases, most often because the defense refused to agree to treatment. Thus, our results are less easy to generalize to larger groups of batterers than the results of the Palmer and Feder experiments. On the other hand, because all batterers included in our sample had to have agreed to treatment, our study presumably did not include batterers who were unmotivated. Of course, all participants were court-mandated; they did not volunteer for treatment of their own volition. Still, it is common knowledge in Brooklyn Criminal Court that misdemeanor batterer defendants are not facing jail time, and participants in treatment certainly knew from counsel that they were choosing the batterer program over another alternative to incarceration. The point about motivation is key, since it has often been argued (see, for example, Rosenfeld, 1992) that treatment cannot be expected to work for individuals there against their will.

The difference between our study and others in *how* batterers were mandated to treatment also has implications for comparison groups. The Palmer and Feder studies compared probationers assigned to treatment to probationers who had similar supervision conditions except for the treatment mandate. In other words, treatment was compared to the absence of treatment. In contrast, our work compares batterers assigned to treatment to batterers assigned to a community service program irrelevant to the problem of violence. The comparison between batterer treatment and an irrelevant treatment is appropriate for judicially-mandated treatment referrals (since all convicted batterers must receive *some* sentence), just as the treatment/absence of treatment comparison is appropriate for probation-mandated referrals.

The Palmer experiment found a significant effect of treatment although the sample size was surprisingly small because the treatment effect size was extraordinarily large. Our work planned sample size based upon an examination of effect sizes described in the literature. Thus, the design contains sufficient power into to provide for adequate tests of the effects of treatment upon several indicators of violence and attitudes.

Due to fortuitous circumstances, we wound up splitting our treatment sample into two subsamples distinguished by density of treatment sessions. All batterers randomly assigned to treatment were mandated to attend 39 hours of psycho educational group treatment based upon the Duluth model. However, some batterers received the 39 hours in 26 weekly sessions while others received it in longer biweekly sessions for 8 weeks. The former treatment model maximized time that batterers remained in treatment while the latter reduced the chances that batterers' initial motivation would flag over time.

Finally, our work included both short-term (6-month post-sentence) and intermediate-term (12-month post-sentence) follow-up on treatment outcomes. Short-term outcomes are important to assess because any effects of treatment may be short-lasting. We know that the likelihood of violence declines as time passes from the time a domestic complaint is made to the police (see, for example, Davis and Taylor, 1997). Any early differences in violence due to treatment might therefore disappear as violence in the control group came down over time. Longer term follow-up is also important to determine whether any short-term effects of treatment hold up in the months after batterers are no longer attending treatment and under court control.

II. METHOD

Overview

The study was conducted using a true experimental design in which 376 criminal court defendants were mandated to attend a 40-hour batterer treatment program or to complete 40 hours of community service. The random assignment was made at sentencing, after all parties (judge, prosecutor, and defense) had agreed to batterer treatment, if it was available based on the random assignment process.

Batterers and victims were interviewed about new violence on three occasions: At the time of sentencing, six months after sentencing, and twelve months after sentencing. Official data on new complaints to the police and new arrests were gathered six and twelve months after sentencing.

Cases Included in the Study

The sampling frame consisted of spousal assault cases in Kings County (New York) Criminal Court in which all parties had agreed in principal to accept batterer treatment, if the defendant was accepted by the Alternatives to Violence (ATV) program. This proved to be a small percentage of cases adjudicated within the course of intake. Intake began on 2/19/95 and ran through 3/1/96. During that time, 376 cases were taken into the sample, about 1-1/2 cases per day. During the same period, roughly xxx??? domestic violence cases were adjudicated (i.e., had dispositions other than dismissal), or about yyy??? per day.

In nearly two-thirds (64%) of the cases in the study, defendants were charged with 3rd degree assault (a class A misdemeanor). An additional 19% were charged with felonious assault (although pleas would be to misdemeanor charges). The remaining 17% were charged with violating restraining orders, menacing, harassment, and other charges. Court dispositions on cases in the sample were most commonly guilty pleas followed by a conditional discharge (68% of the sample) or probation (8% of the sample). Twenty-three percent of the cases were adjourned in contemplation of dismissal (a form of pretrial diversion in which cases are dismissed and records expunged if defendants avoid arrest and adhere to judicial conditions for six months). Conditional discharges and probation place defendants under court control for a period of one year, compared to a period of six months for most adjournments in contemplation of dismissal.

Batterers were all males with a median age of 31 years. The sample contained a plurality (36%) of African-Americans, with substantial numbers of men from Latino (28%) and West Indian (21%) origins as well. Sixty-two percent had graduated high school and just 4% had graduated college. Only about half (54%) of the men reported being employed full time, and just 40% had been continuously employed during the past year. Roughly one-third (36%) reported household income under $10,000/year, while 26% earned between $10,000 and $20,000, and 37% $20,000 or more.

Victims all were females with a median age of 29 years. Six in ten victims (59%) were black, 30% were Latino, and 9% white.[1] The proportion of victims who graduated high school (66%) was comparable to the proportion of high school graduates among batterers reported above. Fewer victims, however, were employed (38%) and a large proportion (43%) received public assistance. Surprisingly, just 9% of the victims reported the batterer as their primary source of assistance. Victims were poorer than batterers, with close to half (46%) reporting household incomes of under $10,000/year.

Victims and batterers had been together a median length of time of 5-1/2 years. On average, violence had begun occurring by two years into the relationship. About two-thirds of victims and defendants lived together at the time of arrest (70% according to batterer interviews/ 62% according to victim interviews). Most batterers in the sample were in current romantic relationships with the victims either as legal spouses (37% according to batterers/ 33% according to victims), live-in boyfriends (19% according to

---

Victim racial profiles differ from defendant ethnic profiles reported above because the questions were asked somewhat differently on the respective interviews. The proportion of victims categorized as "black" corresponds closely to the proportions of defendants categorized as "African-American" (36%) plus the proportion categorized as "west-Indian" (21%).

batterers/ 11% according to victims), or live-out boyfriends (9% according to batterers/ 6% according to victims). Victims and batterers were no longer in a current relationship (33% according to batterers/ 49% according to victims). A large majority of batterers had children in common with the victim (63% according to batterers/ 79% according to victims).

Sixty-two percent of victims said that they had called the police in the past because of their perpetrator's abuse. Forty-eight percent of the victims had filed a police complaint against their perpetrator in the past. Thirty-four percent of the victims had an order of protection against their perpetrator in the past. Twenty-three percent of the victims stated that the perpetrators had been arrested in the past for abusing them. According to official records, 39% of batterers had been arrested previously for any type of crime.

Treatments

There are two ways to conceive of a control treatment for assessing the effects of batterer treatment programs. One is to compare batterer treatment to the absence of treatment. For example, when batterer treatment is left by judges to the discretion of probation officers, assignment to treatment or no treatment can be made at the time of probation intake. This is the method being used in Feder's current study for NIJ.

That option was not available to us since, in New York City, probation for misdemeanor spouse abuse charges is very rare: Judges are the ones who mandate batterers to treatment, and completion of the program is normally the only condition of plea arrangements. It clearly was not possible to suggest to criminal justice officials that they let selected defendants simply walk with no sanctions. Therefore, we needed an alternative sanction for the control group -- a sanction which was irrelevant to the battering problem that resulted in the men's arrest. Community service, as defined below, was such a sanction and criminal justice officials agreed to use it as an alternative to ATV for men designated by researchers as controls. All participants in our experiment were assigned to receive either 40 hours of group batterer treatment or 40 hours of community service.

Batterer treatment The batterer treatment program was Victim Services' Alternatives to Violence (ATV), based upon the Duluth model. The original model mandated 26 weeks of attendance at a weekly group meeting that lasted one hour. The course was rooted in a feminist perspective and assumed that domestic violence is a by-product of male and female sex roles which result in an imbalance of power. The 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, participants were encouraged to take responsibility for their anger, actions, and reactions. Sessions were conducted in either English and Spanish by two leaders, one male and one female.

ATV had changed its format just at the time that the experiment began, expanding the number of required hours from 1-1/2 hours once a week for 12 weeks to 1-1/2 hours once a week for 26 weeks. The change was made to conform with New York State guidelines and was in line with national trends. However, the lengthened program became a sore spot for Legal Aid Society attorneys who defend the vast majority of defendants in Brooklyn Criminal Court judged to be indigent. While Legal Aid administrators had pledged cooperation (and, indeed, made good on that pledge), staff attorneys began to advise their clients against involvement in the new version of the ATV program. Intake slowed to the point that we would have been unable to complete intake within any reasonable time frame. At a meeting with Legal Aid staff attorneys we realized that their objections to ATV stemmed from the increased time that their clients were under court control and from the increased session fees that their clients paid over the course of 26 sessions.

It became clear that, if we were to complete intake, we would have to accommodate the Legal Aid attorneys' objections to the 26-week batterer treatment program. Therefore, with the help of ATV administrators, we designed a new 8-week format through which participants could complete the same 40 hours of group time through bi-weekly 2-1/2 hour sessions with lower fees per session. The new format began to be offered after the first 129 participants had been assigned to 26-week groups. From 8/15/95 until the end of intake, defendants were offered a choice between 8-week and 26-week formats. In practice, no one chose the 26-week option once the 8-

week groups became available. Thus, the final 61 ATV participants were assigned to the 8-week groups.

Community service Defendants rejected by lottery from batterer treatment were mandated by judges to participate in 70 hours of community service. Typically, the service was performed over a two-week period. For offenders who were employed, flexible hours were arranged over a two-month period in order that they could continue their jobs. Participants were assigned to work on renovating housing units, clearing vacant lots to make way for community gardens, painting senior citizen centers, and cleaning up playgrounds -- all activities which would not be expected to impact on abusive behavior. In the course of their service, participants were given education about drugs and HIV. Interested individuals were also referred to drug, HIV, or employment counseling programs.

Participants in both batterer treatment and community service programs were expelled from the programs if a pattern of non-attendance developed (for ATV, three misses constituted grounds for dismissal from the program). For the men assigned to batterer treatment, such cases were referred to the prosecutor's office for action. At the discretion of the district attorney's office, delinquent cases were returned to the court calendar and new sentences could be imposed. In practice, few cases were actually restored to the calendar because the period of court supervision typically was drawing to a close by the time a clear pattern of non-compliance was established and a restoral request was completed.

Follow-up on delinquents was more reliable for the community service group. The organization running that program had the ability to place cases of delinquents on the court calendar themselves, rather than recommending to the prosecutor that cases be restored. If the court issued an arrest warrant for non-compliance, the community service program had enforcement staff who executed the warrants.

Assignment Process and Case Intake

Cases were drawn from three of eight post-arraignment parts in Kings County Criminal Court. Two of the parts were specialized domestic violence parts. The third was the jury trial part where domestic violence and other cases were transferred if a negotiated disposition could not be reached. At the point at which judge, prosecutor, and defense had reached agreement on batterer treatment as an appropriate disposition, the prosecutor called the ATV office in the court building. Either the ATV intake person or a research assistant picked up the defendant in court and brought him to the ATV office for an intake interview.

Upon completion of the interview, the defendant's name and case identifier were entered onto the next line of a logbook. Each line of the book had a pre-assigned treatment designation (batterer treatment or community service) determined through the use of a random number table. The use of the log with pre-determined treatment assignments and the presence of a research assistant on the three busiest days of the week helped to ensure the integrity of the random assignment process. Defendants assigned to batterer treatment were given a start date (usually within a week of intake) and directions to the class.

The defendant was accompanied back to the courtroom and the prosecutor informed of the lottery assignment. The prosecutor informed the judge who then accepted a disposition consistent with the assignment. In 28% of control cases judges overrode the lottery decision to deny batterer treatment and mandated the ATV program for defendants who had been assigned to community service. There were no judicial overrides of cases randomly assigned to the ATV program.

Follow-Up Measures and Rationale

The literature suggests that batterer treatment is designed to reduce violence against women by changing batterers' cognitive understanding about the roles of men and women in society and in relationships. Programs also aim to change batterers' attitudes toward the legitimacy of using violence against family members and to teach batterers ways to resolve interpersonal conflicts without resorting to violence.

Because the most important outcome of treatment is reduction of violence, we included several measures of new violence in victim-batterer relationships. The violence measures (described more fully below) were: new arrests; new crime reports (which may or may not result in an arrest); and self-reports of violence by victims and batterers. These same indicators have become commonly-used in studies which track households where domestic violence occurs, for example, in NIJ's SARP research (see, for example, Fagan, Garner, and Maxwell, 199??). The three violence indicators do not always behave in similar ways (see, for example, Davis and Taylor, 1997), so it is important to capture a variety. Each of the violence measures was captured at 6 and 12 months after the time that batterers were sentenced. Victim and batterer self-reports were obtained through (primarily) telephone interviews. Crime report and arrest data were obtained from official records.

In addition to capturing information on new violent acts, the interviews also assessed attitudinal and cognitive behaviors among batterers and victims. For both groups we measured attitudes toward violence in the family and conflict resolution skills. We also measured for both batterers and victims whether their cognitive styles tended toward internal or external locus of control. That is, did they believe that they could influence events or did they believe that things happened to them? It seemed plausible that, if batterer treatment succeeded in engendering in batterers a greater sense of responsibility for their actions, they would become more internal on locus of control. Finally, the interview schedules included for victims only measures of psychological adjustment. If treatment of the batterer led to changes in the way that they acted toward their partners then, we believed, that women's self-esteem and sense of well-being might improve.

Interview Methodology

We attempted interviews with defendants and victims on three occasions: (a) at case intake (date of court disposition), (b) six months after intake, and © twelve months after intake. Interviews with batterers were conducted in person in the court building just prior to assigning them to either batterer treatment or community service. In subsequent interviews with batterers and all interviews with victims, telephone was the modality of choice. Because we considered the victim interviews more accurate than batterer interviews for assessing new violence, we put special efforts into interviewing victims. When telephone attempts failed, we sent teams of interviewers to victims' homes. If the home interview attempts also failed, we mailed letters offering first $25 and then $50 for completion of an interview. In the third interview wave for victims we turned over 70 difficult cases to a licensed private investigator as a last resort. The private investigator used available computer databases to track victims who had moved and provide us with current addresses. He did not confront victims or their acquaintances, and interviews for women he located were conducted by our staff over the phone. Ultimately, this additional tracking methodology added virtually nothing to the interview success rate.

Completion rates Our completion rate with victims was 50% for the first interview, 46% for the second interview, and 50% for the third interview. First interviews with batterers were obtained with 95% of the sample because interviews were obtained when defendants were present at intake in court for the treatment program. Subsequent completion rates were 40% for the second interview and 24% for the third interview. The fact that attrition among victim interviews was substantially lower than among batterers results from the extra lengths (incentives, in-person visits) to which we went in order to obtain the victim interviews.

The refusal rate for both victims and batterers was quite low (7% and 13%, respectively). The primary reason for not completing interviews with victims and batterers was inaccurate or outdated information obtained from prosecutor files. We had a core group of 23% of victims whom we were unable to contact on any of the three interview occasions. In many of these cases, we found out definitively that the victims had moved, and we suspect that this was the case with most of this group. We have found in research in other cities as well (Davis, Smith, and Nickles, 1997) that court-involved domestic violence victims are a highly transient population with marginal attachment to addresses. Many of those staying with the batterer or with family members at the time of arrest move within a short period of time thereafter.

Interview completion rates did not vary significantly by treatment. Batterer completion rates for experimentals and controls were 94% and 96% at time 1; 42% and 38% at time 2; and 28% and 20% at time 3. Victim completion rates for experimentals and controls were 51% and 50% at time 1; 41% and 50% at time 2; and 52% and 48% at time 3.

Interview rates did vary, however when broken down by some case characteristics. We examined variation in victim and batterer interview completion rates according to batterer age, education, income, employment status, ethnicity and prior arrests. (We used batterer rather than victim characteristics because the former were available for virtually the entire sample and because batterer characteristics have been the primary control variables used in other research on interventions to prevent domestic violence.) In addition, we examined variation in victim and batterer interview rates according to whether the parties were involved in a current, versus an ex-, romantic relationship. We uncovered no significant differences in interview completion rates for either victims or batterers as a function of batterer age, income, employment status, education, prior arrests, or nature of victim/batterer relationship. Neither was there a significant difference in batterer interview completion according to ethnicity. However, ethnicity was correlated with completion of victim interviews: Interviews were completed with victims in 62% of the cases in which batterers were black compared to 76% of the cases in which batterers were non-black (the vast majority of these were non-black latinos).[2]

Interview content Measures on victim and batterer interviews included (a) background information (violence histories and demography); (b) measures of new violence; © beliefs about domestic

[2]Chi-square = 7.99, $p < .01$.

violence; (d) conflict management skills; and (e) locus of control. In addition, victims were administered a short scale measuring well-being. Interviews at the three time points were identical except for the omission of background information on second and third interviews.

A) <u>Background information</u>: (1st interview only)

We assessed violence history in the current relationship between victim and batterer and violence outside of the current relationship perpetrated by batterers and experienced by victims. We also collected limited demographic data (age, ethnicity, marital status, socio-economic status).

B) <u>Measures of recidivism</u>:

To assess frequency and severity of violence, we employed Harrell's (1991) adaptation of the Conflict Tactics Scale (Straus, 1979). Harrell's scale measures the frequency of a range of 11 different violent acts.

The reference period for the scale was the previous two months (as opposed to the previous six months for the criminal justice measures). We reasoned that, if treatment did make a difference, it would take some time to have its effect. Thus, asking victims to report at the six month interval about the entire period would inevitably include violent incidents committed shortly after cases were assigned to treatment. The two month reference period we decided upon ensured that any violence reported would have occurred after batterers had been in treatment for a good length of time.

C) Beliefs about domestic violence

Part of the treatment program curriculum was to encourage batterers to recognize the rights of women not to be abused and to reevaluate the rights of men to use violence to control women. To measure generalized beliefs of batterer and victim about the legitimacy of spouse assault, we used a scale based on the "Inventory of Beliefs about Wife Beating Scale" (Saunders, Lynch, Grayson and Linz, 1987). We began pretesting using the Saunders, et. al. scale intact. However, we soon discovered that many items had little variation. That is, batterers overwhelmingly endorsed the socially desirable choices. These items were dropped and others added, making up a new scale of ten items.

D) Conflict management strategies

We assessed conflict resolution skills of victims and batterers using Harrell's (1991) measure of Conflict Resolution Skills. Harrell's scale is loosely based on Form N of the Straus Conflict Tactics Scale.

E) Locus of control

To assess the degree to which victims and batterers perceived outcomes as contingent upon their actions, we originally attempted to employ Rotter's (1966) Internal-External Locus of Control Scale. However, in pretesting, we discovered prevalent comprehension problems with the Rotter scale. Therefore, we drew 12 items from the 40-item Nowicki-Strickland Internal-External Control Scale (Nowicki and Duke, 1974). This scale is an adaptation of the Children's Nowicki-Strickland I-E Scale, and is thought to be less difficult than Rotter's scale. The items selected were those that seemed most relevant to spouse abuse (e.g., "Do you feel that most of the time it just doesn't pay to try hard because things never work out right anyway?" or "Most of the time do you find it hard to change a friend's mind?").

F) Well-Being (Victims only)

To measure well-being of victims, we used the Life Satisfaction (Index B)(Neugartin, Havighurst, and Tobin, 1961). The scale contains 12 items, each with three ordered response options.

G) Self-esteem

We used the Rosenberg Self-Esteem Scale:(Rosenberg, 1979) to gauge self-perceptions of victims. This 13-item scale asks individuals to rate their extent of agreement (from strongly agree to strongly disagree) with a series of statements about themselves, such as "I am able to do things as well as most other people."

Information Collected from Criminal Justice Records

Computerized records of the Criminal Justice Agency (CJA) and of the New York City Police Department (NYPD) were searched to determine if the batterer was arrested for a new crime or if a new crime report was filed during the study period. CJA's database of New York City arrests was accessed via the court docket numbers of cases in or sample. Docket numbers led us to defendant NYSID (state criminal identification) numbers, which we used to determine if the defendant had had subsequent arrests during the 12 months since sentencing on the sampled case. (All CJA record checks covered at least 12 months, and some covered as many as 26 months.) When new cases were found, the arrest date and charge were recorded. In addition, the docket number was used to search the district attorney's computer database to determine whether the victim in the new case was the same as the victim in the original. Because the searches were conducted using ID numbers, we are confident that our information on new arrests is highly accurate.

The computerized records of the NYPD were searched to determine whether new crime complaints had been filed against the defendant since sentencing in the original case. These searches, conducted by NYPD personnel, were conducted using batterer names and incident addresses. Therefore they were subject to errors in spelling of batterer names or street names in address checks. Also, each police precinct maintains its own database. When batterers commit a crime outside of their home precinct, their home precinct is supposed to receive a record, but we do not know how reliably information is transferred across precinct boundaries. When hits (new incidents) were found, officers recorded the dates of new incidents, the nature of the complaint, and whether the complaint involved the same victim as the original case. As a result of these shortcomings, we expect that the NYPD data undercounted violence reported to the police. We have no reason to believe that the extent of undercounting would vary according to experimental treatment.

We combined the CJA and police data into one measure of new criminal justice involvement in the form of arrests or crime complaints. This parallels the method used by Maxwell (1998) in the most recent reanalysis of data from NIJ's SARP experiment.

III. TREATMENT EFFECTS

Analysis Plan

Our initial decision in data analysis was whether to analyze according to the original two-group design, or to capitalize on the fact that we actually had three treatment groups (8-week, 26-week, control). We examined the data both ways, and discovered that there were substantial differences in outcomes between the two different lengths of batterer treatment. Therefore, we have chosen to present the data broken down into three-group comparisons. However, the same analyses reported here were conducted as well using two-group comparisons with essentially the same pattern of differences between control and treatment groups.

Our initial design called for examining treatment effects six months after sentencing. This interval was chosen to coincide roughly with the end of the 26-week program for subjects assigned to the batterer treatment condition. We reasoned that any treatment effects would be maximal after subjects received the full treatment "dosage". However, effects might decay with the passage of time after program completion. This could happen either because the men assigned to batterer treatment became more violent as time since program completion increased or because control subjects became less violent as more time passed since the incident that led to their arrest.

During the course of our investigation, we were fortunate to receive additional funding from NIJ to enable us to follow subjects up to one year post-sentence. This allowed us to determine if any effects of batterer treatment that were observed immediately upon completion held up over time. Accordingly, we have divided our analyses into short-term (through 6 months post-after assignment to treatment) and long-term (through 12 months after assignment to treatment) effects.

Comparisons Evaluations of batterer treatment pose a challenge for researchers in part because many of those who start treatment programs do not finish them. This was true for our sample as well (see section below on attendance). The temptation in such instances is to compare only those who complete treatment (and therefore get the full "dosage") to a comparison group. However, we followed the example of the SARP investigators in our decision to analyze cases according to the treatment to which they were assigned rather than according to the treatment that they received. This is the course most frequently recommended in both the criminal justice literature and medical literature on clinical trials, although "crossovers" result in loss of statistical power when "analyzing as randomized" (Weinstein and Levin, 1989). However, there are two compelling arguments for our approach.

First, the alternative (analyzing cases according to the actual treatment they receive) runs a serious risk of defeating the purpose of randomizing in the first place, i.e. creating groups of cases equivalent prior to treatment. In our case, the crossovers were created because judges intervened in the random assignment process. Their abrogation of the random assignment in a minority of cases clearly was not a random process. Therefore, it is likely that including such cases in the "treated" group would obviate the initial equivalence that we had sought through randomization. A second argument for analyzing as randomized was made by Gartin (1995). He argues that, in policy studies such as ours, the issue is not the effect of the treatment per se, but the effect of a *policy to apply treatment.*

Sherman (1992) proposes following the "analyze as randomized" dictum as long as the proportion of treatment crossovers does not exceed the proportion of cases with negative outcomes. Our study has a 14% crossover rate due to judicial overrides of random assignments to the control group. This is slightly higher than the one year combined arrest rate of 11% (our most conservative outcome measure), but below the one year combined crime report rate of 17% and the one year victim reports of 19%.

Appendix A presents a comparison of characteristics and violence outcomes for the judicial overrides versus the rest of the controls. None of the differences approached statistical significance, although it must be kept in mind that the number of override cases was small (n=52). There was a substantial percentage difference between override and other controls in the proportion arrested for new crimes against the victim within 12 months of sentencing (21% versus 12%, respectively, p=.14). Assuming that treatment reduces violence, the effect of our analyze as randomized strategy is to reduce the magnitude of treatment effects, i.e. to make the statistical tests more conservative and rejection of the null hypothesis less likely.

Statistical tests At each of the two time points we conducted identical sets of analyses. We began by examining two measures of *prevalence*: (a) new *criminal justice incidents* involving the same victim and (b) new reports of violence made by victims during the course of research interviews. The basic prevalence comparisons between the experimental conditions were done as simple bivariate comparisons.

The prevalence tests were followed by two additional tests on each measure at both time points. The first test was either a Poisson or a negative binomial regression, testing whether the distribution of failures (i.e., cases in which new violence occurred) differed according to treatment. Poisson and negative binomial regression were developed specifically for the kind of distribution of failures that we observed, i.e. a large majority 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 analysis of variance, even with log or other transformations of the data.

The second test was proportional hazard analysis, examining differences between treatment conditions in elapsed time to first failure. In other words, even if no differences were observed between treatments in the proportion or frequency of new violent incidents, it is still possible that one group failed earlier than another.

Finally, for each of the two time points, we examined treatment effects upon measures of cognitive changes, including conflict resolution skills, beliefs about domestic violence, and locus of control. These tests were performed using analysis of variance.

Introduction of covariates In the negative binomial and proportional hazard tests, we added to the model a set of covariates in addition to the treatment variable. The introduction of covariates in analyzing data from a randomized experiment is unusual and, strictly speaking, is not necessary: Randomization ought to ensure that other measures, known or unknown, that are related to the failure measure, such as the suspect's age or prior criminal record, are similarly distributed across the treatment groups and therefore will not bias the basic experimental treatment comparisons described above. In our case, this goal seems largely to have been achieved (see section on pre-treatment comparisons between groups in the last chapter).

However, introducing covariates is increasingly common even in analyzing data from randomized experiments (Patel, 1996; Armitage, 1996). There are several reasons why this is the case. First, statistical controls for other factors tend to improve the precision of the treatment comparisons and correct for any major imbalance in the distribution of these measures across treatments that may have occurred by chance (Armitage, 1996). Second, since the suspects assigned randomly to the same treatment group are not exactly alike, statistical controls can address the natural variations between suspects within each treatment group (Gelber & Zelen, 1986). Third, while an experimental analysis typically tests for only the average effects of treatment across all subjects, whatever their characteristics, additional nonexperimental hypotheses can specify other expected direct effects, like age on the outcomes, and how treatment effects may vary across dimensions of other uncontrolled extraneous factors such as marital status, employment level or prior criminal records.

The tests for the additional direct effects will follow the models that test only for the direct effect of the treatment on the outcome of interest. The nonexperimental measures included the defendant age, ethnicity, employment level, prior arrests, and relationship status with the victim. All of these measures have in some fashion been shown in prior research as predictors of general offending patterns (Blumstein, Cohen, Roth & Visher, 1986; Gottfredson & Gottfredson, 1988), as well as violence between intimates (Fagan & Browne, 1994; Fagan, Garner & Maxwell, 1997; Hotaling and Sugarman, 1990). That is, it increases the chances of finding a treatment effect if one exists. In our analysis, we used a set of covariates which included defendant age, employment status, race, marital status, and prior arrests.

Adding covariates to the analysis also allows us to specify two sets of interaction terms. These interactions will model how two measures of social control (marriage and employment) may mediate the direct effect of the treatment on intimate aggression.

The choice of marriage and employment as the tested mediators is based upon a review of research in other areas of domestic violence interventions that had found these particular measures of social control as important factors in understanding how treatments may not necessary work equally for all batterers (see Sherman, Smith, Schmidt & Rogan, 1992; Pate & Hamilton, 1992). There are numerous ways of testing for the interaction of two independent measures on the outcome measures. Following Hardy (Hardy, 1993), the interaction terms will be specified in such a way that they represented the product of two independent measures that were both coded as dummy (0 and 1) variables. The product or the new interaction term also had the values of 0 or 1, with the suspects taking the value of 1 if they also had 1 on both of the other two measures, and 0 when they had a value of 0 for either or both of the other two measures.

Correcting for missing information Much if not all research in behavioral, economic, and social science is plagued with problems of missing information (Berk, 1983; Weisberg, 1985; Dubin & Rivers, 1989; Winship & Mare, 1992; Little & Schenker, 1995; Breen, 1996). In general there are two types or causes of missing information; item nonresponse and unit nonresponse (Little & Schenker, 1995). In the former type, missing information takes the form of unobserved or unmeasured information on one or more variables for a small subset of cases in a database. This sort of missing information can indicate systematic differences in subjects within the sample that if ignored can lead to a less efficient estimate of an effect size or the complete withdrawal of certain cases from the sample in specific analyses (Weisberg, 1985; Little & Schenker, 1995).

The second type of missing information occurs when cases included in a study represent nonrandom samples of a population. This type of missing data is often referred to as sample selection bias or unit nonresponse. Unlike the first type of missing data (item nonresponse) which is often due to researchers not recording certain responses, this type of missing data is typically created when subjects act in a manner that makes it impossible for the researchers to obtain responses from them for many if not all of the survey's questions (Dubin & Rivers, 1989). The non-respondents' actions may include such things as not listing their telephone numbers, which would exclude them from studies that use telephone numbers as the means for sampling, or being unemployed, which would exclude them from studies that can only sample from those employed. A person's decision not to have a telephone listed or not to look for a job may represent a random process, but it could also be nonrandom. The nonrandom selection of cases from the entire population into a study is itself a social process and an aspect of social science that is often overlooked (Winship & Mare, 1992). For this project there were two opportunities for sample selection bias to occur, one opportunity was at the six-month victim interview and then again at the twelve month interview. Both of these selection opportunities will be independently addressed using separate selection models.

The following analysis addresses systematically both of the missing information problems. In the case of the missing information among some cases on the nonexperimental covariates, a two-step process suggested by Weisberg (1985) was followed to replace the missing data with valid information. The first step was to locate an alternative source of data for the measures with missing information and to use these alternative sources to replace the missing data in the primary database with valid information. After most of the missing data was replaced with valid data from an alternative source, we then moved onto the second step which was to use a statistical technique of imputing quasi-valid values for the remaining missing data. For this particular project we replaced the missing data using a multiple regression imputation procedure. This step specifically involved constructing a regression model that computed 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.

In regards to the second problem of sample selection bias or missing victim interviews, a two step process proposed by Heckman (1979) was employed. 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 latent measure. This step required two different models, one for the six month interview and one for the twelve-month interview. The two latent measures were then entered into the original substantive outcome models as independent measures (one for each interview period) to more fully specify the structured relationship between the dependent and the set of independent measures.

Treatment Attendance Rates

We first compared differences in attendance rates between the 8-week and 26-week groups. We expected that attendance would be better when treatment was compressed into a shorter span in the 8-week groups.

The results, displayed in Table 5, were far more pronounced than we could have guessed. Roughly similar proportions of batterers began treatment in the 8-week and 26-week groups. Seventy-seven percent of those assigned to the 8-week groups attended at least one class compared to 71% of those assigned to the 26-week groups. But graduation rates were dramatically different. Sixty-seven percent of the men assigned to the 8-week groups graduated compared to just 27% of those assigned to the 26-week groups.[3] We conclude that a much larger proportion of those assigned to treatment were exposed to the full treatment in the 8-week groups than in the 26-week groups.

Criminal Justice Incidents

Simple prevalence of new criminal incidents involving the same

[3] Chi-square (1)= 27.72, $p < .001$.

Table 5: Attendance in 8 vs 26 week batterers' group			
	No attendance	**Some attendance**	**Graduated**
26-week format (n=129)	29%	44%	27%
8-week format (n=61)	23%	10%	67%

victim at six months and 12 months after assignment to treatment (i.e., date of sentencing) is presented in Table 6. At six months, reports to the police of new violence involving same victim and perpetrator differed significantly between treatment groups. Seven percent of the 26-week group failed at six months according to this measure compared to 15% of the 8-week group and 22% of the control group. A similar pattern is evident in Table 2 for criminal justice incidents 12 months after assignment. At 12 months, 10% of the 26-week group failed. The 8-week treatment and control groups are virtually indistinguishable, with failure rates of 25% and 26%, respectively.

We now turn our attention to multivariate tests of criminal justice incidents. Multivariate models include Poisson regression models of the rate of offenses and Cox regression models of time to first new criminal justice incident. Both sets of models utilize all of the data captured in record searches which, for most cases, goes well beyond twelve months post-assignment. Record checks were done after the last sampled case had reached 12 months post-assignment, so longer follow-up times were available for most cases. (With Cox regression, longer follow-up times increase statistical power.)

Poisson regression of annual rate of criminal justice incidents Typically OLS regression is used when the quantity of a dependent measure of interest is specified rather than the quality or presence of some event. However, the application of OLS in the instances where the specified dependent variable is the

Table 6: Prevalence of criminal justice incidents involving same victim and perpetrator.

	6 months after assignment*	12 months after assignment**
26-week batterer treatment (n=129)	7%	10%
8-week batterer treatment (n=61)	15%	25%
Control (community service) (n=186)	22%	26%

* Chi-square (2)=12.35, p.003

**. Chi-Square (2)=13.13, p.001

count or a rate of some event is problematic (Gardner, Mulvey & Shaw, 1995). Overall, there are two reasons why OLS is inappropriate: (1) the OLS estimations can produce negative predicted values; and, (2) the hypothesis test used in OLS assumes certain properties of the variance of scores that are unlikely to be met with count data. To address these two problems, the Poisson and related regression models have over the last twenty years begun to replace the OLS regression as the primary means of analyzing dependent measures that are based on a counting process (Land, McCall & Nagin, 1996).

The Poisson regression specifically models in a multivariate context the number of events during an interval of time. Generally, the observed distribution of the counts of events takes on a Poisson like curve, which is one where the number of cases per increasing count is less than the previous count. Because this sort of distribution is useful for handling infrequent events (e.g., instances where most cases have a value of zero or one), the Poisson regression has become invaluable to criminologists. Due to this property, our analysis used Poisson regression when addressing the question of whether the treatment reduced the quantity or rate of failures found in the officially recorded database of new domestic incidents reported to the authorities.

We adjusted the count of criminal justice incidents to an annualized rate to account for the unequal follow-up time across the suspects. This count includes all known recorded offenses that took place after the treatment assignment and makes no distinction or adjustment for the severity or type of criminal offense. The first regression model (Model One; Table 7) provides the results from the classic experimental analysis: the mean of the dependent measure disaggregated by three treatment groups (control group, short treatment, and long treatment). This first regression shows that only the long treatment group had a significant reduction (40%) in the average number of new offenses when compared with the control group. The difference between the control group and short treatment group was not significant, but the direction of the treatment effect was also negative. Model Two then builds on Model One by adding additional control measures to account for the natural heterogeneity between and within the three experimental comparison groups. Again, the long treatment group shows a significant reduction in the number of offenses compared the control group. Beyond this one significant experiment-treatment-effect, no other control measures show either a significant increase or decrease in the number of officially recorded offenses. This lack of a significant effect includes the measures of the suspect's age, ethnicity, employment status, and prior arrest, which have all been found in other research has significant predictors of recidivism among domestic violence batterers (Maxwell, 1998). However, beside the null effect for age, all of the effects from the other four control measures are in their expected direction.

Models Three and Four add two sets of treatment by control interaction to the independent measures regressed in Model Two.

Table 7

Poisson Regression of Annual Rate of Any Officially Recorded Offense

	Model 1			Model 2			Model 3			Model 3		
Model Parameters	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)
ATV												
Short	-0.24	0.30	0.8	-0.24	0.29	0.8	-0.28	0.46	0.8	0.02	0.35	1.0
Long	-0.58	0.24	0.6 *	-0.57	0.24	0.6 *	-0.30	0.34	0.7	-0.90	0.36	0.4 *
Age				0.00	0.01	1.0	0.00	0.01	1.0	0.00	0.01	1.0
Ethnicity(African-American)												
Hispanic				-0.29	0.25	0.7	-0.28	0.25	0.8	-0.28	0.25	0.8
West Indian/Caribbean				-0.47	0.30	0.6	-0.47	0.30	0.6	-0.45	0.30	0.6
Other Race				-0.33	0.32	0.7	-0.32	0.32	0.7	-0.31	0.32	0.7
Married				0.19	0.20	1.2	0.22	0.20	1.2	0.14	0.26	1.1
Employed				-0.24	0.21	0.8	-0.12	0.26	0.9	-0.28	0.21	0.8
Prior Arrest				0.35	0.20	1.4	0.36	0.20	1.4	0.38	0.20	1.5
ATV * Employment												
Short * Employment							0.07	0.58	1.1			
Long * Employment							-0.52	0.49	0.6			
ATV * Married												
Short * Married										-0.65	0.60	0.5
Long * Married										0.66	0.49	1.9
Intercept	-1.10	0.13	*	-1.08	0.43	*	-1.17	0.44	**	-1.03	0.43	*
Model Fit												
Log likelihood	-241.71			-236.52			-235.88			-234.52		
Restricted Log likelihood	-244.89			-244.89			-244.89			-244.89		
Chi-square	6.36			16.74			18.02			20.75		
P-value	0.04			0.05			0.08			0.04		

First, Model Three provides the results for the two treatment (short and long) by employment interaction terms. This model shows that neither of the two treatment groups were either more or less effective at reducing offenses among those employed versus those not employed. Finally, Model Four provides the results for the treatment by marital status interaction terms. This final regression model shows that those victims and suspect who were not married have likely accounted for the significant direct treatment effect, as the suspects in the not married but in the long treatment group were the only ones with a reduced number of officially recorded offenses.

Time-to-first criminal justice incident To examine time from case assignment to first new incident reported to the police, we used Cox regression. The Cox regression model or the semiparametric proportional hazard 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). This analytical model has become an important tool for researchers evaluating criminal justice-based programs, since it can account for the uneven follow-up periods that are characteristics of therapeutic treatment or correctional intervention programs. In other words, this model can accurately analyze data collected on subjects over a time that is not equal in length nor indeterminable.

The Cox model specifically involves constructing a base-line hazard function for the event of interest (e.g., new arrest, new drug use or any other discernable transition) that is dependant only those cases that are uncensored at a particular time-period. This hazard function is then defined as the probability of failing during any particular time interval (e.g., a day), if the individual has survived to the start of that interval (Lee, 1992). The model can then introduce one or more prognostic variables which are used to estimate whether the baseline hazard function is dependent on the level of each independent measure while jointly controlling for the effects of the other endogenous variables. In addition, time-covariant factors can likewise be introduced to test whether the baseline hazard function is dependent on a particular time interval or is proportional overtime. This report will capitalize on the Cox regression model when the analysis of the officially recorded data is concerned with the question of whether treatment influenced the likelihood that aggression had occurred again by the suspects against the victims which were also known to the police.

Table 8 provides the results from five regression models of the hazard or time-to-first new officially recorded offenses. Again, Model One provides the classic experimental analysis. This first model, similar to the one reported in Table 7, shows that the odds of a new offense were significant reduced among the long treatment group compared with the control group. In other words, at any particular time after the treatment assignment the likelihood of the first new offenses was reduced about 50 percent among those in the long treatment group when they were compared with the control. Model Two also shows that this effect is likely not variant overtime as the two time-covariant by treatment interaction terms are not significant. Model Three then builds on Model One by adding the control measures used in the earlier regression model (Table 8; Model Two), and drops the time-covariance terms because they were not significant. Again, the direct negative effect of the long treatment remained significant. However, unlike the earlier regression model, two control measures are now significant and in their effects are in the expected direction. First, the "other" racial group had a significant reduced likelihood of failing when compared to the African-Americans. Besides this significant effect, those with a prior arrest had a significant increased risk of failing at anytime during the follow-up period.

The final two regression models reported in Table 8 provide the results from the same two sets of interaction terms that were reported on earlier in regards to the number of failures. Here the interactions are testing whether the hazard rates for the treatment comparison are dependent on whether the suspect and victim were married or whether the suspect was employed. Both of the regression models suggest that the direct negative effect of treatment is likely mediated by both the marital and employment status of the suspect. More specifically, those suspects assigned to the long treatment who were also not married or not employed had a longer average period of survival than those married or employed. In other words, marriage and employment increased the risk of

Table 8

Cox Regression Model of Time-to-first New Officially Recorded Offenses Against Same Victim

Model Parameters	Model 1 b	Model 1 s.e.	Model 1 Exp(B)	Model 2 b	Model 2 s.e.	Model 2 Exp(B)	Model 3 b	Model 3 s.e.	Model 3 Exp(B)	Model 4 b	Model 4 s.e.	Model 4 Exp(B)	Model 5 b	Model 5 s.e.	Model 5 Exp(B)
ATV															
Short	-0.21	0.29	0.8	-0.52	0.64	0.6	-0.15	0.30	0.9	-0.16	0.47	0.9	0.108	0.36	1.1
Long	-0.72	0.26	0.5 **	-1.36	0.63	0.3 *	-0.74	0.26	0.5 **	-0.75	0.39	0.5 *	-0.96	0.36	0.4 **
ATV															
Short * Time				0.00	0.00	1.0									
Long * Time				0.00	0.00	1.0									
Age							0.01	0.01	1.0	0.01	0.01	1.0	0.01	0.01	1.0
Ethnicity(African-American)															
Hispanic							-0.26	0.26	0.8	-0.26	0.26	0.8	-0.27	0.26	0.8
West Indian/Caribbean							-0.50	0.31	0.6	-0.50	0.31	0.6	-0.50	0.31	0.6
Other Race							-0.76	0.39	0.5 **	-0.76	0.39	0.5 *	-0.74	0.39	0.5 *
Married							0.09	0.22	1.1	0.09	0.22	1.1	0.05	0.26	1.0
Employed							-0.28	0.22	0.8	-0.27	0.26	0.8	-0.32	0.22	0.7
Prior Arrest							0.53	0.22	1.7 **	0.53	0.22	1.7 *	0.55	0.22	1.7 *
ATV * Employment															
Short * Employment										0.01	0.60	1.0			
Long * Employment										0.03	0.53	1.0			
ATV * Married															
Short * Married													-0.65	0.62	0.5
Long * Married													0.50	0.53	1.6
Model Fit															
Log likelihood		1035.47			1035.47			1035.47			1035.47			1035.47	
Restricted Log likelihood		1027.11			1025.58			1010.80			1010.79			1008.15	
Chi-square		7.90			9.15			24.40			24.80			7.903	
P-value		0.02			0.05			0.00			0.01			0.02	

earlier failure among those assigned to the longer treatment group. Nevertheless, the overall effect for the long treatment group was still towards decreasing the risk (see Models 1 & 2), the effect was likely just not equal across all suspects.

Incidents Reported by Victims to Research Interviewers

Simple prevalence of victim reports of violence to research interviewers is reported in Table 9. The table contains victim reports on surveys done approximately six and 12 months after assignment to treatment. In each survey, victims report on the immediately preceding two months. At six months, virtually no differences are apparent between groups. Twenty-three percent of the 26-week group reported a new incident compared to 19% of the 8-week group and 21% of the control group. Differences were larger at 12 months, following the same pattern as the criminal justice incidents, but still did not approach statistical significance. At 12 months, 15% of victims whose cases were assigned to the 26-week group reported a new incident within the past two months compared to 18% of victims whose cases were assigned to the 8-week group and 22% of victims whose cases were assigned to the control group.

Negative binomial regression There is one major limitation of the Poisson regression used above in analyzing treatment differences in new incidents reported to authorities. That is the assumption that the mean and the variance are identical and equal to the expected mean (Land, et al., 1996). When this assumption is not met the model is considered overdispersed, which

Table 9: Prevalence of incidents reported by victims to research interviewers.

	6 months after assignment*	12 months after assignment**
26-week batterer treatment	23% (n=52)	14% (n=66)
8-week batterer treatment	19% (n=26)	18% (n=33)
Control (community service)	21% (n=93)	22% (n=90)

*Chi-Square (2)= 0.15, p=.926
**Chi-Square (2)=1.86, p=.394

can lead to incorrect estimations of variances and misleading inference about the effects of independent measures. To adjust for this problem an additional term that reflects the "unexplained between-subject difference is included in the regression model." (Gardner, et al., 1995, p. 393). This additional term in turn changes the Poisson model into a negative binomial model, which only assumes that the dependent measure looks like a Poisson Distribution, and not that all individuals have the same mean rate. Because the negative binomial model through the addition of one term removes the Poisson's assumption of equity, it provides greater flexibility for accurately representing the relative frequency of observed event count data (Land, et al., 1996). With the victim interview data on reports of violence, we performed a test which showed that ovedispersion was present. Therefore, we substituted a negative binomial for a Poisson model.

Tables 10 and 11 provide the results from both the six and twelve month victim interviews. Here, the outcome measure, extracted from a modified CTS, is delineated as the maximum number of aggressive incidents by the suspect against the victim that she reported happening over the two months preceding the two interviews. The results show after correcting for sample selection bias and adding a term to address overdispersion, that neither the short nor the long term treatments seemed to have reduced significantly the frequency of aggression at about the six or twelve months periods. However, at both time periods and for both treatment groups the direction of the effect is negative (e.g.,

Table 10

Negative Binomial Regression of the Past Two Month Frequency of Victimization @ Six Month Survey

	Model 1			Model 2			Model 3			Model 3		
Model Parameters	b	s.e.	Exp(B)	b.	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)
ATV												
Short	-1.53	1.34	0.2	-1.12	1.43	0.3	0.49	2.72	1.6	-2.93	2.16	0.1
Long	-0.88	0.90	0.4	-1.02	0.91	0.4	-0.74	1.12	0.5	-1.05	1.28	0.3
Age				0.04	0.08	1.0	0.06	0.08	1.1	0.05	0.08	1.0
Ethnicity (African-American)												
Hispanic				1.38	1.11	4.0	1.68	1.25	5.4	1.36	1.09	3.9
West Indian/Caribbean				0.32	1.25	1.4	0.11	1.29	1.1	0.13	1.45	1.1
Other Race				0.96	1.84	2.6	1.04	1.96	2.8	0.66	1.96	1.9
Married				-1.39	1.24	0.2	-1.52	1.27	0.2	-1.74	1.45	0.2
Employed				-0.44	1.02	0.6	0.23	1.312	1.3	-0.36	1.04	0.7
Prior Arrest				1.11	1.14	3.0	0.82	1.14	2.3	1.24	1.12	3.5
ATV * Employment												
Short * Employment							-2.63	3.39	0.1			
Long * Employment							-0.68	1.73	0.5			
ATV * Married												
Short * Married										2.58	2.96	13.1
Long * Married										-0.02	1.92	1.0
Intercept	-4.79	7.63		-7.95	6.94		-10.75	8.94		-7.57	6.81	
Selection Bias ratio	6.45	9.07		8.15	8.38		10.34	10.31		7.48	8.18	
Scalar	11.04	3.30	***	9.0	2.46	***	8.81	2.47	***	8.72	2.40	***
Model Fit												
Log likelihood	-199.2177			-193.87			-192.77			-192.77		
Restricted Log likelihood	-545.7404			-474.46			-473.13			-468.86		
Chi-square	694			561.18			560.41			552.18		
P-value	0.00			0.00			0.00			0.00		

Table 11

Negative Binomial Regression of the Past Two Month Frequency of Victimization @ Twelve Month Survey

Model Parameters	Model 1 b	Model 1 s.e.	Model 1 Exp(B)	Model 2 b	Model 2 s.e.	Model 2 Exp(B)	Model 3 b	Model 3 s.e.	Model 3 Exp(B)	Model 3 b	Model 3 s.e.	Model 3 Exp(B)
ATV												
Short	-0.94	1.01	0.4	-0.79	1.18	0.5	-2.16	2.29	0.1	-2.10	1.66	0.1
Long	-1.29	0.81	0.3	-1.57	1.07	0.2	-1.70	1.47	0.2	-0.95	1.12	0.4
Age				0.02	0.05	1.0	0.01	0.05	1.0	0.01	0.05	1.0
Ethnicity(African-American)												1.0
Hispanic				-0.85	1.06	0.4	-0.57	1.29	0.6	-0.51	1.21	0.6
West Indian/Caribbean				0.34	1.18	1.4	0.43	1.44	1.5	0.51	1.36	1.7
Other Race				0.10	1.69	1.1	0.00	1.85	1.0	-0.02	1.61	1.0
Married				-0.86	1.30	0.4	-0.98	1.28	0.4	-0.51	1.17	0.6
Employed				-0.80	1.18	0.4	-1.18	1.41	0.3	-1.00	1.12	0.4
Prior Arrest				-1.03	0.92	0.4	-0.83	0.96	0.4	-0.90	0.93	0.4
ATV * Employment												1.0
Short * Employment							2.06	2.90	7.8			1.0
Long * Employment							0.20	2.11	1.2			1.0
ATV * Married												1.0
Short * Married										1.90	2.47	6.7
Long * Married										-3.52	2.82	0.0
Intercept	1.62	7.93		4.41	9.14		3.56	9.60		5.49	9.21	
Selection Bias ratio	-1.02	10.12		-3.98	11.51		-2.27	12.12		-5.14	11.49	
Scalar	13.92	3.36	***	11.97	3.34	***	11.65	3.36	***	10.35	2.98	***
Model Fit												
Log likelihood	-191.03			-187.30			-186.44			-182.28		
Restricted Log likelihood	-617.99			-551.84			-546.37			-529.49		
Chi-square	853.92			729.09			719.85			694.43		
P-value	0.00			0.00			0.00			0.00		

reduction in the average frequency of aggression). In regards to the other nonexperimental factors tested, no other variable was a significant predictor of an increase or decrease in the level of aggression as well, and the two sets of interactions terms likewise indicated that the null direct effect of the treatments were not dependent on level of social control. In other words, the two sets of regression models reported in both tables are poor at explaining any variation of the frequency of intimate aggression beyond just the mean.[4]

Cognitive Change Measures

Our measures of the cognitive change in batterers included conflict resolution skills, beliefs about domestic violence, and locus of control. Each of these scales has problems for use as a

[4] A set of identical logistic regressions model were also estimated using a dichotomized dependent measure for aggression instead of count of aggression. The results were similar to the extent that no treatment groups were significant different from the controls. The only noteworthy differences between the two estimation procedures was that the logistic produced a significant positive effect of prior arrest on the odds of failing and the long treatment produced a nonsignificant increase in the odds of failing. Otherwise, the models were very similar

measure of cognitive change in batterers. The beliefs about domestic violence scale has limited reliability statistics. The conflict management strategies scale similarly has been little-studied for test-retest reliability. The locus of control test has been problematic for use with batterers because of it assumes a fairly high level of cognitive functioning. (We sought to mitigate this problem by using a children's version of the scale.) Moreover, it could be argued that batterer intervention groups could teach participants how to answer items "correctly" on any of these scales without any true change in cognitive beliefs or behavior. Still, the group of scales used to assess cognitive change represented the most commonly-used indices at the time the study was conducted.

The original analysis plan called for a repeated measures MANOVA test using the same set of covariates described above in the recidivism analyses. However, we were unable to carry out this plan due to serious limitations in the data. First, the internal validity of the scales was low. The conflict resolution skills scale was respectable, averaging .71 over the six and twelve month interviews with batterers. Reliability of the locus of control scale averaged .69 over the six and 12 month interviews. The beliefs about domestic violence scale had lower reliability, averaging .57 over the two follow-up points. Second, ns for the three cognitive measures were very low. At the 6-month interview, we had 149 cases, and just 88 cases at the 12-month interview.

Means and standard deviations for each of the three tests at each of the two time points are presented in Table 12. For each scale, means across the three treatment groups are remarkably similar, and none of the univariate F-ratios also presented in Table 12 come close to statistical significance. We have, therefore, no basis for claiming that treatment changed batterers' attitudes or ways of dealing with conflict. But again we note that limitations in the scales and in our data do not permit an adequate test of this hypothesis.

Table 12: Means and Standards deviations for psychosocial outcomes*

		Conflict Resolution Skills	Attitudes Toward Spouse Abuse	Internal/External Locus of Control
6-month survey	Control (n=69)	18.1 (6.3)	25.2 (5.5)	3.5 (2.0)
	8-week (n=27)	19.6 (6.1)	25.2 (6.5)	2.9 (2.4)
	26-week (n=53)	18.0 (5.7)	25.2 (5.1)	3.2 (2.1)
		F(2,116)=0.57	*F(2,146)=0.00*	*F(2,146)=0.41*
12-month survey	Control (n=37)	19.3 (6.2)	24.4 (4.1)	3.5 (2.0)
	8-week (n=18)	19.1 (6.0)	25.1 (4.8)	3.1 (2.5)
	26-week (n=33)	19.9 (5.9)	25.9 (4.6)	3.1 (2.1)
		F(2,62)=0.91	*F(2,85)=0.35*	*F(2,85)=0.51*

*Numbers on parentheses are standard deviations.

IV. CONCLUSIONS

Our initial analyses showed that men assigned to a group treatment program for batterers were less likely to be the subject of future crime complaints involving the same parties than men assigned to an irrelevant treatment (community service). This difference was most pronounced at six months after group assignment, but held up over a full year.

Subsequent analyses revealed interesting findings about length of treatment. Due to fortuitous circumstances, we wound up splitting our treatment sample into two subsamples distinguished by density of treatment sessions. All batterers randomly assigned to treatment were mandated to attend 39 hours of psycho educational group treatment based upon the Duluth model. However, some batterers received the 39 hours in 26 weekly sessions while others received it in longer biweekly sessions for 8 weeks. The former treatment model maximized time that batterers remained in treatment while the latter reduced the chances that batterers' initial motivation would flag over time.

Our results showed that far more men successfully completed the 8-week group than the 26-week group. We expected, therefore, that men assigned to the 8-week group would have a lower rate of recidivism than men assigned to the 26-week group. However, only the 26-week group was statistically different from the control group on future crime complaints: The 8-week group and the control group were indistinguishable. Victim reports of violence to research interviewers showed a similar pattern, but differences between treatment conditions did not approach statistical significance.

Batterer intervention can be looked upon in one of two ways. It may be a learning process in which attitudes and behavior are modified in a relatively permanent way, Or it may be that batterer intervention simply suppresses violent behavior for the duration of treatment, but no permanent changes are effected. Our results do not support the model of treatment as a change process: If that were true, then the men in the 8-week group (who were finished with treatment long before the follow-up period was up) ought to have been as non-violent as their 26-week counterparts (who were in treatment for most of the follow-up period). Yet that is not what our results showed. Also, we did not find evidence that treatment altered attitudes toward spouse abuse, further suggesting that there was no basis for permanent changes. (However, the reader is again advised of serious limitations in the cognitive change scales and data.)

Our results, then support the suppression model of batterer intervention. But they are only suggestive since the study was not designed to test the validity of various models of the treatment process. Moreover, they are at odds with other studies which have not tended to find a difference in recidivism according to length of treatment (Edelson and Syers, 1990; Gondolf, 1997a). Many current batterer programs are going to longer treatment models, but there is also substantial pressure from the defense bar and economics to keep time in treatment to a minimum. Thus, the question of whether treatment works only as long as men attend groups is key to intelligent policy formulation.

How do our results fit into the literature on batterer treatment? If we concentrate only on the four quasi- and two true experiments (including ours), then we note that five of the six (Harrell, et. al. is the lone exception) reported results in the expected direction and all reported statistical significance on at least one outcome measure.

But even more striking are the effect sizes (i.e., strength of association between treatment and criminal outcomes) from these investigations. Effect size has been argued to be a more important index of treatment effects than statistical significance (e.g., Cohen, 1992; Rosenthal, 1991). It provides a measure of delectability of an effect which is independent of the baseline rate to which it is being compared (Bem and Honorton, 1994). (The power to detect the difference between .55 and .25 is different from the power to detect the difference between .50 and .20.)

We computed effect sizes for five of the six quasi- and true experiments. (Harrell's anomalous work was omitted from this analysis.) The effect sizes were computed on proportions of repeat violence culled from police records because it was the most commonly available measure from this group of studies. Effect size was assessed using Cohen's *h* (Cohen, 1988). In the five batterer treatment studies that found evidence in favor of treatment, effect sizes ranged from 0.108 to 0.946 (see Table 13). To place these effect sizes in context, consider the effect size of one of the early large clinical trials on the effect of aspirin on heart attack rates. In that research, more than 22,000 subjects were randomly assigned to take aspirin or a placebo. The study was stopped after six years because it was already clear that the aspirin treatment was effective ($p<.00001$) and today it is common medical practice for doctors to prescribe aspirin to prevent second heart attacks. Yet the effect size, as measured by Cohen's h, was only 0.068. Against this standard, the effect sizes seen in batterer treatment studies are quite substantial.

A common technique in meta-analysis is to give studies quality ratings and then correlate the ratings with effect sizes. If the effect size decreases as quality of the research goes up, it is a good indication that the effect is not real (see, for example, Utts, 1991). This has often been the case in criminal justice. For example, early literature on pretrial diversion was generally positive; but when a true experiment was conducted, no effect of diversion upon subsequent criminal behavior was found (Baker and Sadd, 1979).

In contrast, the effects do not seem to disappear in the batterer treatment literature as the studies become more rigorous. Referring to Table 13, it is clear that treatment effects do not decline as we move from quasi-experiments to true experiments. The average effect sizes for the two true experiments (0.412) is virtually identical to the average for the quasi-experiments (0.416).

Table 13:

Effect Sizes

Quasi-Experiments	Recidivism		Effect Size
	Treated	Untreated	
Dutton (1986)	4%	40%	0.946
Chen et al. (1989)	5%	10%	0.193
Dobash et al. (1996)	7%	10%	0.108
Average			0.416

True-Experiments	Recidivism		Effect Size
	Treated	Untreated	
Palmer et al. (1992)	10%	31%	0.537
Davis and Taylor (1997)	5%	13%	0.287
Average			0.412

Taken together, these studies provide a case for rejecting the null hypothesis that treatment has no effect on violent behavior toward spouses. However, the number of useful studies is small and more well-designed studies are warranted before coming to firm conclusions.

Recommendations for Future Research

The evaluations that have been done can provide useful information to future researchers. From these studies, we have estimates of treatment effect size which can be used to determine appropriate sample sizes for future investigations. Researchers will not need to guess whether they need 50 cases or 500 cases in order to attain the requisite statistical power needed to detect real effects.

We recommend that several standards be applied to future investigations into whether treatment has an effect on violence. First, as recommended by Fagan (1996) and others, randomized experiments should be the design of choice. We recognize that random assignment when applied to judicial mandates to treatment are likely to prove difficult or impossible (since it is tantamount to sentencing by lottery and requires the agreement of prosecution, defense, and judiciary). However, true experimental designs are not unrealistic when applied to probationers who are mandated to treatment at the discretion of probation administrators. Jurisdictions in which treatment mandates are at the discretion of the probation agency are prime potential settings for research.

Our study provides a good illustration of the difficulties that can be encountered implementing a true experimental design. We had to make substantial concessions to court officials in order to gain their cooperation. Judges were allowed to override assignments to the control group in exceptional cases. This produced a high rate of judicial overrides of cases assigned to the control group. As we showed in the last chapter, the effect of including the override cases in the control group was to make the tests of treatment effects more conservative. (Yet, we still found large treatment effects.) Also, we had to offer a treatment alternative that was more palatable to the defense than the lengthy and costly version that we started with. This proved to be a fortuitous change, however, since we found substantial differences in outcomes between men assigned to the 8-week and 26-week groups. We agree with the opinion of most serious researchers, however, that the benefits of random assignment outweigh the potential difficulties.

Second, measures and follow-up intervals need to be standardized so that results can be compared across studies. Too many studies have relied only upon batterer self-reports, known to vastly underestimate the true incidence of abuse (for an expanded discussion of this point, see Rosenfeld, 1992). The same kinds of measurement standards used in the National Institute of Justice's Spouse Abuse Replication Project (SARP) studies ought to be applied to batterer treatment: Investigations ought to include victim reports, crime complaints made to the police, and arrests. Batterers ought to be tracked at six-month intervals for at least one year, and preferably two. Short-term measures are needed to assess immediate program effects -- effects that may be transient. Longer-term follow-up is needed to determine whether treatment leads to permanent changes. The use of both short-term and long-term measures is especially important in light of the suggestions from some of the SARP sites that law enforcement intervention may have deterrent effects in the short-run, but facilitating effects on battering in the long run (for a detailed discussion of measurement issues in the SARP data, see Garner, Fagan, and Maxwell, 1995).

Third, investigations of the effects of batterer treatment need to be explicit in defining the standard against which treatment is being evaluated. Too many studies have compared men who go through batterer treatment to men who receive unspecified other sentences in the courts. To gauge the effects of treatment compared to the absence of treatment, it is imperative that batterers in the control group receive nothing relevant to reducing their propensity to batter. This may be possible when using a sample of probationers, some of whom are assigned to batterer treatment in addition to regular supervision and others of whom are assigned only to normal supervision regimes.

Fourth, researchers need to find ways to minimize attrition from treatment programs. Batterer program attrition typically runs greater than half of all participants assigned to treatment. This poses a serious dilemma for researchers, who must choose between analyzing groups as assigned (that is compare all individuals assigned to treatment to all individuals in the control condition) and comparing only program completers to controls. If treatment attrition is high, the first alternative results in overly conservative estimates of program effects since the treatment group is made up of many individuals exposed to minimal or no treatment. On the other hand, comparing only treatment completers to controls biases the analysis in favor of finding significant treatment effects since those who complete treatment are the "cream" of the group of batterers assigned to treatment.

Sherman (1992) argues that, assuming treatment attrition can be minimized, the clear preference is to "analyze as randomized". The critical question, according to Sherman, is whether the proportion of cases treated differently than the random assignment is larger than the proportion of cases with negative outcomes. On the hand, analyzing according to treatments as assigned becomes a problem when the treatment often fails to be delivered. A high rate of treatment "crossovers" reduces statistical power and increases the likelihood that an effective treatment will go undetected (Gartin, 1995; Weinstein and Levin, 1989).

The best way out of this dilemma is to minimize treatment crossovers, most commonly attributable to treatment program attrition. Suggestions are that treatment attrition can be minimized by telescoping treatment into a short time span and by imposing penalties for failure to attend classes. Also, studying treatment programs located within corrections institutions -- where batterers have no choice about attending sessions -- would provide a way around the attrition problem. Such an institutional setting would provide a vehicle to examine the "dosage-response curve" indicating how treatment outcomes vary according to the number of sessions batterers are exposed to. (So-called "dosage-response curves" confound treatment effects with differences in participant motivation when they are based on the number of sessions batterers voluntarily attend.) This issue is important in light of the trend toward longer treatment programs yet -- excepting the present results -- unsubstantiated by empirical findings indicating that lengthy programs work better than shorter ones.

Fifth, researchers ought to make explicit issues which may restrict the extent to which their findings can be generalized. Particular attention needs to be given to the sample of batterers who participate in a research study. Are they court-mandated? Do they have extensive prior criminal histories or not? Do defendants have a chance to volunteer for treatment or are they sent to treatment regardless of their willingness to participate? Also potentially important is the criminal justice context within which treatment studies are set. Treatment program effectiveness may vary according to local court practices, linkages between agencies, sanctions for non-compliance, and so forth.

Finally, researchers need to find ways to maximize interview response rates when interviewing victims about continuing abusive behavior from their spouses. It is common to have interview success rates below 50% when contacting victims six months or later. There are good reasons why rates are so low: Researchers are interviewing victims who did not initially agree to participate, they must rely on inaccurate contact information from the files of criminal justice agencies, and domestic violence victims and offenders are notoriously transient. Nonetheless, with interview success rates below 50%, it is difficult to make the case that interview data are representative of the sample as a whole. However, with sophisticated methods of follow-up and judicious use of financial incentives, it should be possible to attain relatively respectable response rates (see Sullivan, Rumptz, Campbell, Eby, and Davidson, 1996 for a discussion of minimizing survey attrition with battered women samples).

There are parallels between the batterer treatment literature today and the literature on the rehabilitation of criminal offenders 15 or so years ago. In both literatures, the problem is not too few studies, but a paucity of sophisticated research. Calls that were made years ago by the National Academy of Sciences (Martin, Sechrest, and Redner, 1981) for agreement on outcome measures and randomized experiments in rehabilitation are just as relevant today for batterer treatment. The evolution in sophistication of batterer treatment studies is encouraging. Using randomized experiments and other designs that have a high degree of internal validity, we soon should be able to say whether batterer treatment works and to specify which program models are most effective.

REFERENCES

Adams, D. (1988). Counseling men who batter: A profeminist analysis of five treatment models. In M. Bograd & K. Yllo (Eds.), *Feminist perspectives on wife abuse* (pp. 177-198). Beverly Hills, CA: Sage.

Armitage, P. (1996). The design and analysis of clinical trials. In S. Ghosh & C.R. Rao (Eds.) *Handbook of statistics, vol. 13: Design and analysis of experiments.* North-Holland.

Baker, S. & Sadd, S. (1979). *Court employment project final report.* New York: Vera Institute.

Bem, D.J. & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomolous process of information transfer. *Psychological Bulletin, 115*, 4-18.

Berk, R. A. 1983. An introduction to sample selection bias in sociological data. *American Sociological Review* 48(3, June):386-98. June.

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

Brannen, S.J. & Rubin, A. (1996). Comparing the effectiveness of gender-specific and couples groups in a court-mandated spouse abuse treatment program. *Research on Social Work Practice, 6*, 405-424.

Breen, R. 1996. *Regression models: censored, sample-selected, or truncated data.* Sage University Papers Series: Quantitative application in the social science. Thousand Oaks: CA: Sage Publiation.

Buzawa, E., & Buzawa, C. (1996) *Domestic violence: The criminal justice response* (2nd edition). Newbury Park: Sage Publications.

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

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*, 98-101.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crowell, N., & Burgess, A. W. (Eds.). (1996). *Understanding violence against women.* Washington, DC: National Academy Press.

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

Davis, R.C. & Taylor, B.G. (In press). Does batterer treatment reduce violence? A synthesis of the

literature.
Women and Criminal Justice, in press

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

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

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

Dubin, J. A., Rivers, D. 1989. Selection bias in linear regression, logit and probit model. *Sociological Methods and Research* 18(2 & 3, November):360-90. November.

Dunford, F. W. (1997). *History of the San Diego project and baseline data, the San Diego Navy Project.* Working draft, University of Colorado.

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

Dutton, D. G. (1988). *The domestic assault of women: Psychological and criminal justice perspectives.* Boston, MA: Allyn & Bacon.

Dutton, D. G. (1995). *The domestic assault of women: Psychological and criminal justice perspectives* (rev. ed.). Vancouver: UBC Press.

Edleson, J. L., & Syers, M. (1990). Relative effectiveness of group treatments for men who batter. *Social Work Research and Abstracts, 26*(2), 10-17.

Eisikovits, Z. C. & Edelson, J. L. (1989). Intervening with men who batter: A critical review of the literature. *Social Service Review, 37*, 384-414.

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

Fagan, J. (1996). The criminalization of domestic violence: Promises and limits. *NIJ Research Report* (January). Washington, DC: National Institute of Justice, U.S. Department of Justice.

Fagan, J. (1989). Cessation of Family Violence: Deterrence and dissuasion. In L. Ohlin & M. Tonry (Eds.), *Family Violence.* Chicago: University of Chicago Press.

Fagan, J., Browne, A. 1994. Violence against spouses and intimates. Panel on the Understanding and Control of Violent Behavior, Committee on Law and Justice, Commission on Behavioral and Social Science and Education, National Research Council. In *Understanding and controlling violence*, ed. A. J. Reiss, J. A. Roth, vol. 3. Washington, D.C.: National Academy Press.

Fagan, J., Friedman, E., Wexler, S. & Lewis, V.L. (1984). *National Family Violence Evaluation: Final Report. Volume 1: ExecutiveSummary and Analytic Findings.* San Francisco: URSA Institute.

Fagan, J., Garner, J., Maxwell, C. D. 1997. *Reducing Injuries to Woman in Domestic Assaults.* Final Report. National Center for Injury Control and Prevention, Centers for Disease Control and Prevention, Department of Public Health and Human Services.

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

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

Ganley, A. (1987). Perpetrators of domestic violence: An overview of counseling the court-mandated client. In D.J. Sonkin (Ed.), *Domestic violence on trial: Psychological and legal dimensions of family violence* (pp. 155-173). New York: Springer.

Gardner, W., Mulvey, E. P., Shaw, E. C. 1995. Regression analysis of counts and rates: Poisson, overdispersed. *Psychological Bulletin* 118(3):392-404.

Garner, J., Fagan, J., & Maxwell, C. (1995). Published findings from the spouse assault replication program: A critical review. *Journal of Quantitative Criminology, 11*, 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, 32*, 425-445.

Gelber, R. D., Zelen, M. 1986. Planning and reporting of clinical trials. In *Medical oncology*, ed. P. Calabresi, P. S. Schein, S. A. Rosenberg, pp. 406-25. New York, NY: Macmillian Publishing Company.

Goldkamp, J.S. (1996). *The role of drug and alcohol abuse in domestic violence and its treatment: Dade county's domestic violence court experiment* (Final Report). Philadelphia,PA: Crime and Justice Research Institute.

Gondolf, E. (1997a). *Multi-site evaluation of batterer intervention systems: A summary of preliminary findings*. Indiana, PA: Mid-Atlantic Addiction Training Institute.

Gondolf, E. (1997b). Batterer typers based on the MCMI: A less than promising picture. Unpublished paper.

Gondolf, E. (1995). *Batterer intervention: What we know and need to know*. Paper presented at the National of Institute of Justice Violence Against Women Strategic Planning Meeting, Washington, DC.

Gondolf, E. (1991). A victim-based assessment of court-mandated counseling for batterers. *Criminal Justice Review, 16* (2), 214-226.

Gottfredson, M. R., Gottfredson, D. M. 1988. *Decision making in criminal justice: toward the rational exercise of discretion.* Ed. J. Feinber, T. Hirschi, B. Sales, D. Walker. Law, Society and Policy. New York: Plenum Press.

Gottman, J. M., Jacobson, N. S., Rushe, R. H., Shortt, J. W., Babcock, J., La Taillade, J. J., & Waltz, J. (1995), The relationship between heart rate reactivity, emotionally aggressive behavior, and general violence in batterers. *Journal of Family Psychology, 9*(3), 227-248.

Grusznski, R. J., & Carillo, T. P. (1988). Who completes batterer's treatment groups? An empirical investigation. *Journal of Family Violence, 3*, 141-150.

Hamberger, L. K,. & Hastings, J. E. (1988). Skills training for treatment of spouse abusers: An outcome study. *Journal of Family Violence, 3*, 121-130.

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

Hamberger, L. K., & Hastings, J. E. (1990). Recidivism following spouse abuse abatement counseling: Treatment and program implications. *Violence and Victims, 5*, 157-170.

Hamberger, L. K., & Hastings, J. E. (1993). Court-mandated treatment of men who assault their partners: Issues, controversies, andoutcomes. In N. Z. Hilton (Ed.), *Legal responses to wife assault: Current trends and evaluation.* Newbury Park, CA: Sage.

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

Hardy, M. A. 1993. *Regression with dummy variables.* Sage University Paper series on Quantitative applications in the social sciences. Newbury Park, CA: Sage Publications.

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

Harrell, A. V., Roehl, J. A., & Kapsak, K. A. (1988). *Family violence intervention demonstration programs evaluation, volume II: Case studies.* Report submitted to the Bureau of Justice Assistance. Washington, DC: The Institute of Social Analysis.

Harris, R., Savage, S., Jones, T., & Brooke, W. (1988). A comparison of treatments for abusive men and their partners within afamily-service agency. *Canadian Journal of Community Mental Health, 7*(2), 147-155.

Healey, K., Smith, C., & O'Sullivan, C. (1997). *Batterer intervention: Program approahes and criminal justice strategies.* Report of Abt Associates to the National Institute of Justice, Washington, DC.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1, January):153-61.

January.

Holtzworth-Munroe, A., Staurt, G. L. 1994. Typologies of male batterers: three subtypes and the deference among them. *Psychological Bulletin* 116(3):476-97.

Hotaling, G. T., Surgarman, D. B. 1990. A risk maker analysis of assaulted wives. *Journal of Family Violence* 5(1):1-13.

Jacobson, N. S., Gottman, J. M., Shortt, J. W. (1995). The distinction between type 1 and type 2 batterers--further considerations: Reply to Ornduff et al. (1995), Margolin et al. (1995), and Walker (1995). *Journal of Family Psychology, 9*(3), 272-279.

Land, K., C., McCall, P. L., Nagin, D. S. 1996. A comparison of Poisson, negative binomial, and semiparametic mixed regression models. *Sociological Methods and Research* 24(4, May):387-442. May.

Lee, E. T. 1992. *Statistical methods for survival data analysis.* Wiley series in probability and mathematical statistics. Applied probability and statistics. New York, NY: John Wiley & Sons.

Little, R. J. A., Schenker, N. 1995. Missing data. In *Handbook of Statistical Modeling for the Social and Behavior Science*, ed. G. Arminger, C. C. Clogg, M. E. Sobel, pp. 39-76. New York, NY: Plenum Press.

Maiuro, R.D., Cahn, T.S., Vitaliano, P.P. & Zegree, J.B. (1987, August) Treatment for domestically violent men: Outcome and follow-up data. Paper presented at the meeting of the American Psychological Association, New York.

Martin, S., Sechrest, L., & Redner, R. (Eds.) (1981). *New directions in the rehabilitation of criminal offenders.* Washington, D.C.: National Academy of Sciences Press.

Maxwell, C. D. 1998. The specific deterrent effect of arrest on aggression between intimates and spouses. diss. Newark, New Jersey: Rutgers, the State University of New Jersey.

Palmer, S. E., Brown, R. A., & Barrera, M. E. (1992). Group treatment program for abusive usbands:
Long-term evaluation. *American Journal of Orthrpsychiatry, 62*(2), 276-283.

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

Pate, A., Hamilton, E. E. 1992. Formal and informal deterrents to domestic violence. *American Sociological Review* 57(October):691-97. October.

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

Rosenbaum, A., & O'Leary, K. (1986). The treatment of marital violence. In N. S. Jacobsen & A. S.
Gurman (Eds.), *Clinical handbook of marital therapy.* NY: Guilford.

Rosenfeld, B. D. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review, 12,* 05-226.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage.

Sampson, R.J. & Laub, J. (1990). Crime and deviance over the life course: The salience of adult social
bonds. *American Sociological Review, 55,* 609-627.

Saunders, D. G. (1996a). Interventions for men who batter: Do we know what works. *Psychotherapy*
in Practice, 2 (3), 81-93.

Saunders, D. G. (1996b). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and Victims.*

Saunders, D. G., & Azar, S. (1989). Family violence treatment programs: Descriptions and evaluation.
In L. Ohlin & M. Tonry (Eds.), *Family violence: Crime and justice, a review of research* (pp. 481-546). Chicago, IL: University of Chicago Press.

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

Sherman, L. W., Smith, D. A., Schmidt, J. D., Rogan, D. P. 1992. Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review* 57(October):680-90. October.

Stark, E., Flitcraft, A. 1988. Violence among intimates: an epidemiological review. In *Handbook of family violence*, ed. V. B. Hasselt, R. L. Morrison, A. S. Bellack, M. Hersen, pp. 293-318. New York, NY: Plenum Press.

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

Toby, J. (1957). Social disorganization and stake in conformity: Complimentary factors in the predatory behavior of hoodlums. *Journal of Criminal Law, Criminology, and Police Science*, *48*, 12-17.

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

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

Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, *6*, 363-378.

Weinstein, G.S. & Levin, B.L. (1989). Effect of crossover on the statistical power of randomized studies. *Annals of Thoracic Surgery*, *48*, 490-95.

Weisberg, S. 1985. *Applied linear regression.* Wiley series in probability and mathematical statistics. Applied probability and statistics. New York, NY: John Wiley & Sons.

Winship, C., Mare, R. D. 1992. Models for sample selection bias. In *Annual Review of Sociology*, pp. 327-50. Palo Alto, CA: Annual Reviews Inc.

APPENDIX A

ITEM FREQUENCIES ON ABUSE SCALE ADOPTED FROM HARRELL (1991)

Item	6-Month Interview (n = 171)	12-Month Interview (n = 189)
1. Forced sex	5%	4%
2. Choked/strangled	3%	3%
3. Threatened to kill	13%	7%
4. Beat up	7%	4%
5. Threatened with weapon	3%	2%
6. Used weapon	2%	1%
7. Threw object	5%	5%
8. Pushed/grabbed/shoved	13%	11%
9. Slapped/spanked	4%	6%
10. Kicked/bit/punched	5%	3%
11. Hit	4%	5%
Any of above	22%	19%

APPENDIX B

DIFFERENCES IN CASE CHARACTERISTICS PRIOR TO TREATMENT

(8- AND 26-WEEK BATTERER TREATMENT GROUPS AND CONTROLS)

	8-week Group (n=61)	26-week Group (n=129)	Controls (n=186)	p
Defendant/Case Characteristics				
Has prior arrests? (% yes)	43%	41%	37%	.66
Batterer employed? (% yes)	67%	63%	64%	.84
Batterer high school grad? (% yes)	64%	64%	61%	.80
Batterer African-American? (% yes)	36%	29%	41%	.04
Batterer age (years)	30.9	33.3	33.2	.17
Batterer/victim married? (% yes)	43%	42%	40%	.89

APPENDIX C

CHARACTERISTICS OF JUDICIAL OVERRIDES AND OTHER CONTROL CASES

	Overrides (n=52)	Other Controls (n=134)	p	
Defendant/Case Characteristics				
Has prior arrests? (% yes)	40%	36%	.65	
Batterer employed? (% yes)	55%	68%	.10	
Batterer high school grad? (% yes)	57%	63%	.80	
Batterer African-American? (% yes)	35%	42%	.77	
Batterer age (years)	33.8	33.0	.58	
Batterer/victim married? (% yes)	51%	37%	.09	
12-MonthRecidivism Outcomes				
Official reports/arrests (% yes)	21%	12%	.14	(183)
Victim reports to interviewers (% yes)	17%	25%	.37	(90)

Does Batterer Treatment Reduce Violence?

A Randomized Experiment in Brooklyn

EXECUTIVE SUMMARY

Robert C. Davis

Bruce G. Taylor

Christopher D. Maxwell

Victim Services Research
346 Broadway, Suite 206
NY, NY 10013

January 2, 2000

ABSTRACT

During the past two decades, pro-arrest laws have resulted in an increasing number of prosecutions of men who assault spouses or girlfriends. Researchers and practitioners have documented the difficulty of altering the behavior of convicted spouse abusers. As the courts have searched for effective sanctions for spouse abusers, they have increasingly come to rely on group treatment programs as the sentence of choice for the widening pool of men convicted of spousal assault.

The greater reliance on batterer treatment programs makes it important that we can document that such programs effectively reduce the propensity of offenders to commit new violence. There is no shortage of evaluations of batterer treatment programs: Some three dozen have appeared in the literature since the 1980s. Most of these studies have methodological deficiencies, which make it difficult to interpret their findings. But evaluation studies have become more sophisticated as time has passed.

The present study represents one of the first attempts to conduct a test of batterer treatment using a true experimental design. The design randomly assigned 376 court-mandated batterers to batterer treatment or to a treatment irrelevant to the battering problem (community service). All men assigned to batterer treatment were mandated to 39 hours of class time. But some were assigned to complete the treatment in 26 weeks and others in eight weeks. Men assigned to the control condition were sentenced to forty hours of community service. For all cases in the study, interviews were attempted with victims and batterers at 6 months and 12 months after the sentence date. In addition, records of criminal justice agencies were checked to determine if new crime reports or arrests had occurred involving the same defendant and victim.

The results showed that treatment completion rates were higher for the eight-week group than for the 26-week group. However, only defendants assigned to the 26-week group showed significantly lower recidivism at 6 and 12 months post-sentencing compared to defendants assigned to the control condition. The groups did not differ significantly at either 6 or 12 months in terms of new incidents reported by victims to research interviewers. We interpret the results to indicate that batterer intervention has a significant effect on suppressing violent behavior while batterers are under court control, but may not produce

INTRODUCTION

Over the past two decades, the law enforcement response to domestic violence has become increasingly tough. 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 domestic violence 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 group treatment programs as the sanction of choice.

There are compelling reasons why group treatment programs for batterers have become a popular mode of court sanction. Even in serious battering cases, many victims choose to stay with abusive partners. Such victims are interested in sanctions which offer them safety from violence, not retribution or punishment that will jeopardize their partner's ability to earn a living. Alternative sanctions commonly used in other crimes have little face validity in abuse cases: There is little reason to believe that fines, community service or probation without special conditions will stop batterers from abusing their spouses.

There is no shortage of evaluations of batterer treatment programs. But the vast majority has serious methodological flaws which make it impossible to distinguish between treatment effects, temporal effects, and selection effects. Generally, the evaluation literature shows an evolution toward more rigorous science since the first batterer treatment studies appeared in the literature in the early 1980s. The study we describe represents one of the first attempts to conduct a test of batterer treatment using a true experimental design which randomly assigns court-mandated batterers to batterer treatment or to a control condition.

The Nature of Batterer Treatment

The first group programs for batterers were begun during the late 1970s. Feminists, victim advocates, and others realized that providing services to victims of abuse and then returning them to the same home environment did little to solve abuse problems (Healey, Smith, and O'Sullivan, 1997). Group treatment was believed to be more appropriate than individual counseling or marital therapy because it expanded the social networks of batterers to include peers who are supportive of being nonabusive (Crowell and Burgess, 1996). Groups also proved to be less expensive than one-on-one counseling sessions. The earliest batterer groups were educational groups which sought to promote an anti-sexist message (Gondolf, 1995). With the passage of time, they gradually incorporated cognitive/behavioral therapeutic techniques and skill-building exercises.

As states introduced pro-arrest statutes during the 1980s the number of batterers arrested and convicted increased, and group treatment became the treatment of choice for the courts. Court-mandated batterer treatment significantly increased and diversified the number of batterer programs nationally (Feazell, Mayers, & Deschner, 1984). A recent estimate places the proportion of court mandates in treatment programs at 80% (Healey, et. al. 1997).

Batterer treatment may be required by criminal courts as part of a pre-trial diversion program, may be ordered by judges as part of a sentence, or may be imposed by probation agencies empowered to set special conditions of probation (Hamberger & Hastings, 1993). In at least one major urban jurisdiction, the district attorney sometimes agrees not to file charges at all if a brief treatment program is completed (Davis and Smith, 1997). In some states (see Ganley, 1987), civil courts as well as criminal may mandate a batterer to treatment (e.g., as a condition related to child visitation).

Many batterer programs are run by probation departments, while others are run by mental health practitioners, family service organizations, or victim service programs. Intake practices vary, with some programs accepting all court referrals and others exercising discretion in excluding persons with prior convictions or substance abuse problems. Supervision of batterers in treatment can most often falls to probation officers, but is sometimes undertaken by others - and increasingly by judges. Historically, supervision has been lax, drop out rates high, and sanctions unevenly applied. Recently, however, supervision has become stricter and sanctions for failure to attend sessions more common.

The Evaluation Literature

Over the last two decades there have been many empirical studies on batterer treatment programs. There are at least six published reviews of over 35 published single-site evaluations (e.g., Eisikovits & Edleson, 1989; Gondolf, 1991,1995; Rosenfield, 1992; Saunders, 1996a; Tolman & Bennett, 1990) and eight research reviews (e.g., Davis and Taylor, in press; Hamberger & Hastings, 1993; Crowell & Burgess, 1996; Dobash, Dobash, Cavanagh & Lewis, 1995; Dutton, 1988, 1995; Rosenbaum & O'Leary, 1986; Saunders & Azar, 1989; Tolman & Edleson, 1995).

However, the volume of the literature is deceptive. In fact, there have been only a handful of investigations that can make any legitimate claims about differences between treated batterers and untreated batterers. The batterer treatment literature has gone through three generations of studies. Most recent have been investigations which have randomly assigned batterers to treatment conditions. These are the strongest designs. Quasi-experiments of varying quality appeared somewhat earlier in the literature. The oldest, and by far the largest, portion of the empirical literature consists of studies which examine only batterers assigned to treatment programs. Included in this set of studies are: (a) studies which assess violence or other individual outcomes only after batterer treatment, (b) studies which measure violence before and after treatment, and © studies which compare violence of batterers who complete treatment with batterers assigned to treatment, but do not attend. Although the methodologies of early studies do not tend to be strong, they are important because they laid the foundation upon which stronger designs could be developed.

Studies Without a Comparison Group

Non-experimental one group post-test only designs

At least 15 published studies have used designs which generate a single measure of treatment effectiveness: violence following completion of treatment (see Table 1). Ten measured recidivism based only upon batterer self-reports. Only four of the fifteen studies had substantial sample sizes (which we have arbitrarily defined as greater than 100) or lengthy follow-up periods (which we have defined as one year or greater).

Recidivism rates in this group of studies vary widely, from 7% to 47% (mean 26%). Interpretation of results is difficult at best without a comparison group or pre-test information with which to compare outcome measures.

Non-experimental one group pre-test and post-test designs

At least seven published studies compared violence among treated batterers after program participation to violence levels prior to participation (see Table 2). Three of the seven studies included both victim and batterer self-reports, but just two had follow-up periods of at least a year and none of the studies examined police records. Two of the seven studies had sample sizes greater than 100. Of the six studies that reported treatment attrition rates, four of the studies had attrition rates of 25% or less.

Table 1: Batterer Treatment Evaluations Using a Post-Test Only Design

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivism	Attrition
Purdy & Nickle (1981)	170	Batterer	6 months	41%	Unknown
Deschner (1984)	12	Batterer	8 months	15%	50%
Feazel, Mayers, and Deschner (1984)	90	Batterer	1 Year	25%	Unknown
Edleson, Miller, Stone, and Chapman (1985)	9	Batterer	7 to 21 weeks	22%	0%
Neidig, Friedman, and Collins (1985)	Unknown	Batterer	4 months	13%	Unknown
Harris (1986)	40	Batterer	2 months to 3 years	27%	Unknown
DeMaris and Jackson (1987)	53	Batterer	20 months	35%	83%
Leong, Coates, and Hoskins (1987)	67	Victim, Police	3 months	19% (Victim) 15% (Police)	76%
Shupe, Stacey & Hazlewood (1987)	148	Victim, Batterer	3 months to a few years	30% (Victim) 18% (Batterer)	31%
Tolman, Beeman, and Mendoza (1987	48	Victim	6 months	47%	68%
Edleson and Grusznski (1988) (Study 2)	86	Victim	9 months	33%	0%
Beninati (1989)	16	Batterer	Unknown	19%	25%
Hamberger and Hastings (1990)	106	Batterer	1 year	30%	16%
Johnson and Kanzler (1990)	687	Batterer	5 months	7%	30%
Tolman and Bhosley (1991)	99	Victim	1 year	42%	50%

Table 2: One Group Pre and Post-Test Design

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivisim	Attrition
Dutton (1986) Part 1	50	Batterer & Victim	6 months to 3 years	Pre-Test 13.4 All DV acts (Batterer reports) / Post-Test 4.6 All DV acts (Batterer reports) Pre-Test 21.3 All DV acts (Victim reports)/ Post-Test 6.1 All DV acts(Victim reports) (For all differences, $P<.05$)	10%
Rosenbaum (1986)	11	Batterer	4 & 6 Months	100% (Pre-treatment) 9% (4 months) 27% (6 Months) ($P< .05$)	18%
Waldo (1986)	23	Batterer	6 Months	Pre-Test 5.1 DV acts / Post-Test 0.29 DV acts ($P < .05$)	Unknown
Shepard (1987)	92	Batterer	14 months	Pre-Test 39% / Post Test 30% (Statistical significance not reported)	25%
Hamberger and Hastings (1988) Part 1	35	Batterer, Victim (Combined measure)	1 year	Pre-Test 20.9 DV acts / Post-Test 5.3 DV acts ($P < .001$)	0%
Meredith & Burns (1990)	125	Batterer, Victim	3 months	Physical, verbal & emotional abuse all reduced at post-test (% not reported)	53%

All seven studies reported lower recidivism rates following treatment (but results of one study were not statistically significant; two studies did not report probability statistics). However, with this type of design, reductions in recidivism cannot be attributed necessarily to the effects of treatment. This is true because studies have repeatedly 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 repeat domestic violence within the next six months after the police intervene (see, for example, Davis and Taylor, 1997; Sherman, 1992; Fagan, Friedman, Wexler, and Lewis 1984). The post-treatment violence rates displayed in Table 2 also average about one-third -- in other words not different than one might expect even if the batterers had not undergone treatment.

Comparing treatment drop-outs versus completers Six studies compared outcomes between batterers who completed treatment and batterers assigned to a treatment program, but who failed to complete treatment (see Table 3). Four of the six studies had sample sizes under 100. Only two of the six studies had follow-up periods of at least one year, and just one included more than a single measure of recidivism.

The most serious flaw in these six studies is that the treated and untreated (dropout) groups are almost certainly not comparable in complex ways prior to treatment. As pointed out by Palmer, Brown, and Barrera (1992), attendance is a confounding factor because better attendance is likely an indication of higher

Table 3: Quasi-Experiment (Dropouts Versus Completers)

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivism
Halpern (1984)	84	Victim	3 months	18% dropouts / 15% completers (N.S.)
Hawkins & Beauvais (1985)	106	Police	6 months	18% Dropouts / 18% completers (N.S.)
Douglas & Perrin (1987)	40	Police	6 months	29% Dropouts / 15% Completers (No Statistics Reported)
Edleson and Grusznski (1988, Study1)	86	Victim	About 5 to 9 months	46% Dropouts / 32% completeres (P < .03)
Edleson and Grusznski (1988, Study 3)	159	Victim	1 year	48% Dropouts / 41% completers (N.S.)
Hamberger and Hastings (1988) Part 2	71	Batterer, Victim, Police (Combined measure)	1 year	47% dropouts / 28% completers (P <.06)

motivation to change, even before treatment. Therefore, differential recidivism between program completers and drop-outs could be due to motivational differences in the two groups that existed prior to treatment. Surprisingly, however, only one of the six studies reported significantly lower recidivism rates for the completers (four of the other five studies were in the predicted direction but either had results that were not statistically significant or did not include inferential statistics).

The best use of this group of studies is to describe the characteristics of people that drop-out of treatment -- information potentially useful to program developers to improve batterer groups. Results have indicated that those who do not complete treatment are more likely to be victims of child abuse (Grusznki & Carrillo, 1988), unemployed (Hamberger & Hastings, 1988;), uneducated (Grusznki & Carrillo, 1988), young (Hamberger & Hastings, 1993), psychologically disturbed (Hamberger & Hastings, 1989; Grusznki & Carrillo, 1988), and substance abusers (Hamberger & Hastings, 1990).

Quasi-Experimental Non-Equivalent Matched Groups

We found four studies in which batterers mandated to treatment by the courts were compared to batterers who received other interventions. This group of studies is the first we have examined which addressed in a rigorous fashion the issue of whether treatment works. There is a notable difference in design details between these four quasi-experiments and the other studies reviewed

Table 4: Quasi-Experiment (Matched Control Group)

Authors of Study	Sample Size	Data Source	Follow-up Time	Recidivism	Attrition
Dutton (1986) Part 2	100	Police	6 months to 3 years	40% No treatment / 4% Treatment (P < .001)	0%
Chen, Bersani, Myers, and Denton (1989)	221	Police	Average of 14 months	10% (0.53 DV acts) No Treatment / 5% (0.35 DV acts) Treatment (P < .05) Perps Attended >75% TX less recidivisim than controls(P<.05)	Unknown
Harrell (1991)	348	Batterer/ Victim (Combined measure), Police	6 months for batterer & victims, 15 and 29 Months for police	15% severe violence No Treatment / 20% Treatment (P=N.S.), 12% physical aggression No TX / 43% Treatment (P<.01) 7% New DV Charges No Treatment / 19% Treatment (P < .05)	24%
Dobash et al (1996)	313	victim & court reports	3 & 12 months	7% treated, 10% untreated (court reports 12 months) 30% treated, 62% Untreated (victim 3 months) 33% treated, 75% untreated (victim 12 months) No probability statistics provided	Unknown

thus far. All four of the studies had sample sizes greater than 100 (see Table 4). None of the studies relied solely on batterer self-reports. All four had follow-up periods of at least one year.

The first quasi-experiment was reported by Dutton (1986). His sample consisted of 100 convicted batterers on probation. He compared 50 batterers who were treated within a cognitive-behavioral group model to 50 batterers who were not designated to receive treatment. The treatment group had a 4% recidivism rate compared to 40% for the control group based upon police reports. However, although Dutton reports that groups did not differ on several demographic measures, pre-treatment comparability of the groups is highly suspect: The control group was composed of batterers whom probation officers did not select for treatment, some of whom were explicitly rejected by therapists as unsuitable for treatment. The treatment group consisted of only batterers who *completed* the treatment program. Dutton does not report what proportion of all batterers assigned to treatment dropped out but, based on other work, we have to assume that it was a large proportion.

Chen et al. (1989) conducted a quasi-experiment involving 120 batterers assigned to treatment by the courts and 101 comparison batterers drawn from court calendars who were not mandated to go to treatment. (No details are given on how the controls were selected or what the outcomes were of their court cases, although the authors state that the samples proved to be well-matched demographically.). Sixty-three percent of the men assigned to

Harrell's analysis included only batterers in the treatment group who actually completed treatment. Comparisons of recidivism were based on a combined measure of the victim and perpetrator reports of violence six months after case disposition. In addition, police records were reviewed 15-29 months after case disposition. Surprisingly, a significantly larger percentage of those in the treatment group committed new violence than those in the control group for two of three measures that she reports. (The third measure is in the same direction, but not statistically significant.). For example, 7% of the control group and 19% of the treatment group were charged with new domestic crimes. While Harrell's study may be limited in its ability to distinguish between selection effects and treatment effects, it certainly adds controversy to the debate about the efficacy of treatment programs.

Recently, Dobash, Emerson-Dobash, Cavanagh and Lewis (1996) reported on a quasi-experiment evaluating a treatment program in Great Britain. Dobash et al. 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. (These differences are reminiscent of pre-treatment differences in Harrell's study.)

It is not specified whether Dobash, et. al. included in their analyses all batterers assigned to treatment, or only those who completed treatment. According to court reports at 12 months follow-up, 7% of the treatment group recidivated compared to 10% of the control group: No statistical tests were reported to indicate whether the difference was significant. 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, although no specific information is provided.) However, the success rate for interviews was low: Dobash et al. interviewed only 43% of the victims at the first follow-up interview, 34% at the second interview, and 25% at the third interview.

Randomized Experiments

As pointed out by Palmer et al. (1992), quasi-experiments on batterer treatment cannot be relied upon to produce unbiased estimates of the effects of treatment. This is true because we cannot know whether batterers assigned to treatment and controls are equivalent prior to application of the treatment. In some quasi-experiments (such as the Dutton, 1986 or Harrell, 1991 studies), we know for certain that selection bias favored finding treatment effects (because the control group was comprised of batterers more prone to recidivate than those in the treated group).

It can be argued that initial differences between groups can be controlled statistically, but this is only true if all relevant initial differences are known to researchers. For example, a researcher may discover pre-treatment differences in employment, marital status, and criminal history between those assigned to batterer treatment and controls, and these differences may be statistically controlled in analyses. However, groups may well have differed on less tangible and more fundamental factors such as emotional maturity as well. If such factors are not controlled (because they are not known) and they are correlated with outcome measures, then the results of the study are uninterpretable. The safest way to ensure that estimates of sample means are unbiased is through random assignment of batterers to treatments.

Palmer et. al. conducted the first experiment with random assignment to a true no treatment control group. The number of subjects in the experiment was far smaller than one would expect to need to detect treatment effects: Fifty-nine probationers were assigned using a "block random" procedure to either a ten-session psychoeducational group (combining group discussion with information) or a no treatment control group: Participants were assigned to treatment if a new group was to commence within three weeks; otherwise they became part of the control group. In only two cases was a defendant assigned to the control condition reassigned by court officials to the treatment condition. Attrition was kept within a respectable range: 70% of the men assigned to treatment attended at least seven of the required 10 sessions.

It is significant that this is one of the only studies to compare all batterers *assigned* to treatment (not just those who completed treatment) with controls. Palmer and her colleagues examined police reports six months post-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 (10% compared to 31%). Even with the small N, this difference was statistically significant. While Palmer et. al. attempted to generate additional violence measures from surveys of interviews and batterers, low response rates combined with a small N precluded any analysis of recidivism based upon interview data.

Two additional randomized experiments are in progress. Dunford (1997) is in the final stages of comparing treatment outcomes for 861 legally married Navy couples in which physical abuse had come to the attention of Navy authorities. These cases were randomly assigned to one of four treatments, including (a) 26-week batterer treatment (based on a cognitive/.behavioral model), (b) 26 weeks of couples counseling, © rigorous monitoring (including monthly calls to victims and semi-annual police record checks), and (d) establishing a safety plan for victims. The safety planning was intended by the investigators as a no-treatment control against which to compare the effects of the other three treatments. (Safety planning was given to victims in each of the other three conditions as well.) This would seem to be a fairly good no-treatment condition, in so far as the men in this condition received no intervention. Victims and batterers are being interviewed every six months over a period of two years. Feder (1996) has assigned batterers placed on probation to either a 26-week educational batterer program based on the Duluth model or a control group not mandated to treatment. Multiple measures of recidivism will be assessed (victim, batterer, police records, probation records) for six months and one year.

Purposes of the Present Study

We sought to add to the incipient literature on randomized studies of batterer treatment. Although any form of design can be criticized, we concur with Fagan (1996) that randomized experiments entail less serious problems than other designs. A properly executed randomization process is the only way to ensure that treatment effects are not confounded with pre-existing subject characteristics. Our study adds to the literature on randomized experiments in several important ways.

Unlike the sites of the Palmer and Feder experiments, batterers in the site of our study were mandated to treatment by judicial order (in the sites of the other two studies, orders to treatment were made by probation departments). This difference has implications for the kinds of batterers studied. The Palmer and Feder studies had a wide sampling frame, including all or most batterers sentenced to probation, regardless of the batterers' willingness or unwillingness to enter into treatment. In our study, batterers were only eligible for inclusion if all parties to the case (prosecution, defense, and judge) agreed that treatment was appropriate. Such agreement was forthcoming in a small percentage of cases, most often because the defense refused to agree to treatment. Thus, our results are less easy to generalize

The Palmer experiment found a significant effect of treatment although the sample size was surprisingly small because the treatment effect size was extraordinarily large. Our work planned sample size based upon an examination of effect sizes described in the literature. Thus, the design contains sufficient power into to provide for adequate tests of the effects of treatment upon several indicators of violence and attitudes.

Due to fortuitous circumstances, we wound up splitting our treatment sample into two subsamples distinguished by density of treatment sessions. (Readers interested in detail on the events that led to the change in treatment length are referred to the full report.) All batterers randomly assigned to treatment were mandated to attend 39 hours of psycho educational group treatment based upon the Duluth model. However, some batterers received the 39 hours in 26 weekly sessions while others received it in longer biweekly sessions for 8 weeks. The former treatment model maximized time that batterers remained in treatment while the latter reduced the chances that batterers' initial motivation would flag over time.

Finally, our work included both short-term (6-month post-sentencing) and intermediate-term (12-month post-sentencing) follow-up on treatment outcomes. Short-term outcomes are important to assess because any effects of treatment may be short-lasting. We know that the likelihood of violence declines as time passes from the time a domestic complaint is made to the police (see, for example, Davis and Taylor, 1997). Any early differences in violence due to treatment might therefore disappear as violence in the control group came down over time. Longer term follow-up is also important to determine whether any short-term effects of treatment hold up in the months after batterers are no longer attending treatment and under court control.

METHOD

Overview

The study was conducted using a true experimental design in which 376 criminal court defendants were mandated to attend a 40-hour batterer treatment program or to complete 40 hours of community service. The random assignment was made at sentencing, after all parties (judge, prosecutor, and defense) had agreed to batterer treatment, if it was available based on the random assignment process.

Batterers and victims were interviewed about new violence on three occasions: At the time of sentencing, six months after sentencing, and twelve months after sentencing. Official data on new complaints to the police and new arrests were gathered six and twelve months after sentencing.

Cases Included in the Study

The sampling frame consisted of spousal assault cases in Kings County (New York) Criminal Court in which all parties had agreed in principal to accept batterer treatment, if the defendant was accepted by the Alternatives to Violence (ATV) program. This proved to be a small percentage of cases adjudicated within the course of intake. Intake began on 2/19/95 and ran through 3/1/96. During that time, 376 cases were taken into the sample.

In nearly two-thirds (64%) of the cases in the study, defendants were charged with 3rd degree assault (a class A misdemeanor). An additional 19% were charged with felonious assault (although pleas would be to misdemeanor charges). The remaining 17% were charged with violating restraining orders, menacing, harassment, and other charges. Court dispositions on cases in the sample were most commonly guilty pleas followed by a conditional discharge Twenty-three percent of the cases were adjourned in contemplation of dismissal (a form of pretrial diversion in which cases are dismissed and records expunged if defendants avoid arrest and adhere to judicial conditions for six months). Conditional discharges and probation place defendants under court control for a period of one year, compared to a period of six months for most adjournments in contemplation of dismissal.

Treatments

Batterer treatment The batterer treatment program was Victim Services' Alternatives to Violence (ATV), based upon the Duluth model. The original model mandated 26 weeks of attendance at a weekly group meeting that lasted one hour. The course was rooted in a feminist perspective and assumed that domestic violence is a by-product of male and female sex roles which result in an imbalance of power. The 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, participants were encouraged to take responsibility for their anger, actions, and reactions. Sessions were conducted in either English and Spanish by two leaders, one male and one female.

ATV had changed its format just at the time that the experiment began, expanding the number of required hours from 1-1/2 hours once a week for 12 weeks to 1-1/2 hours once a week for 26 weeks. The change was made to conform with New York State guidelines and was in line with national trends. However, the lengthened program became a sore spot for Legal Aid Society attorneys who defend the vast majority of defendants in Brooklyn Criminal Court judged to be indigent. While Legal Aid administrators had pledged cooperation (and, indeed, made good on that pledge), staff attorneys began to advise their clients against involvement in the new version of the ATV program. Intake slowed to the point that we would have been unable to complete intake within any reasonable time frame. At a meeting with Legal Aid staff attorneys we realized that their objections to ATV stemmed from the increased time that their clients were under court control and from the increased session fees that their clients paid over the course of 26 sessions.

It became clear that, if we were to complete intake, we would have to accommodate the Legal Aid attorneys' objections to the 26-week batterer treatment program. Therefore, with the help of ATV administrators, we designed a new 8-week format through which participants could complete the same 40 hours of group time through bi-weekly 2-1/2 hour sessions with lower fees per session. The new format began to be offered after the first 129 participants had been assigned to 26-week groups. From 8/15/95 until the end of intake, defendants were offered a choice between 8-week and 26-week formats. In practice, no one chose the 26-week option once the 8-week groups became available. Thus, the final 61 ATV participants were assigned to the 8-week groups.

Community service Defendants rejected by lottery from batterer treatment were mandated by judges to participate in 70 hours of community service. Typically, the service was performed over a two-week period. For offenders who were employed, flexible hours were arranged over a two-month period in order that they could continue their jobs. Participants were assigned to work on renovating housing units, clearing vacant lots to make way for community gardens, painting senior citizen centers, and cleaning up playgrounds -- all activities which would not be expected to impact on abusive behavior. In the course of their service, participants were given education about drugs and HIV. Interested individuals were also referred to drug, HIV, or employment counseling programs.

Participants in both batterer treatment and community service programs were expelled from the programs if a pattern of non-attendance developed (for ATV, three misses constituted grounds for dismissal from the program). For the men assigned to batterer treatment, such cases were referred to the prosecutor's office for action. At the discretion of the district attorney's office, delinquent cases were returned to the court calendar and new sentences could be imposed. In practice, few cases were actually restored to the calendar because the period of court supervision typically was drawing to a close by the time a clear pattern of non-compliance was established and a restoral request was completed.

Follow-up on delinquents was more reliable for the community service group. The organization running that program had the ability to place cases of delinquents on the court calendar themselves, rather than recommending to the prosecutor that cases be restored. If the court issued an arrest warrant for non-compliance, the community service program had enforcement staff who executed the warrants.

<u>Assignment Process and Case Intake</u>

Cases were drawn from three of eight post-arraignment parts in Kings County Criminal Court. Two of the parts were specialized domestic violence parts. The third was the jury trial part where domestic violence and other cases were transferred if a negotiated disposition could not be reached. At the point at which judge, prosecutor, and defense had reached agreement on batterer treatment as an appropriate disposition, defendants were screened by Atv for eligibility and then randomly assigned to batterer treatment or community service. Defendants assigned to batterer treatment were given a start date (usually within a week of intake) and directions to the class.

After assignment to treatment, the defendant was accompanied back to the courtroom and the prosecutor informed of the lottery assignment. The prosecutor informed the judge who then accepted a disposition consistent with the assignment. In 28% of control cases judges overrode the lottery decision to deny batterer treatment and mandated the ATV program for defendants who had been assigned to community service. There were no judicial overrides of cases randomly assigned to the ATV program.

Follow-Up Measures

Because the most important outcome of treatment is reduction of violence, we included several measures of new violence in victim-batterer relationships. The violence measures were: new incidents involving the same victim which were reported to criminal justice authorities and reports by victims of new incidents to research interviewers. These indicators have become commonly-used in studies which track households where domestic violence occurs, for example, in NIJ's SARP research (see, for example, Fagan, Garner, and Maxwell, 1995). Violence indicators do not always behave in similar ways (see, for example, Davis and Taylor, 1997), so it is important to capture more than one. Both measures were captured at 6 and 12 months after the time that batterers were sentenced. Victim self-reports were obtained through (primarily) telephone interviews. Crime report and arrest data were obtained from official records.

In addition to capturing information on new violent acts, the interviews also assessed attitudinal and cognitive behaviors among batterers and victims. For both groups we measured attitudes toward violence in the family and conflict resolution skills. We also measured for both batterers and victims whether their cognitive styles tended toward internal or external locus of control.1 That is, did they believe that they could influence events or did they believe that things happened to them? It seemed plausible that, if batterer treatment succeeded in engendering in batterers a greater sense of responsibility for their actions, they would become more internal on locus of control. Finally, the interview schedules included for victims only measures of psychological adjustment. If treatment of the batterer led to changes in the way that they acted toward their partners then, we believed, that women's self-esteem and sense of well-being might improve.

---

1 Cognitive measures included the Inventory of Beliefs about Wife Beating Scale" (Saunders, Lynch, Grayson and Linz, 1987); Harrell's (1991) measure of Conflict Resolution Skills; and a shortened (12-item) version of the Nowicki-Strickland Internal-External Control Scale (Nowicki and Duke, 1974).

Interview Methodology

We attempted interviews with defendants and victims on three occasions: (a) at case intake (date of court disposition), (b) six months after intake, and (c) twelve months after intake. Interviews with batterers were conducted in person in the court building just prior to assigning them to either batterer treatment or community service. In subsequent interviews with batterers and all interviews with victims, telephone was the modality of choice. Because we considered the victim interviews more accurate than batterer interviews for assessing new violence, we put special efforts into interviewing victims. When telephone attempts failed, we sent teams of interviewers to victims' homes. If the home interview attempts also failed, we mailed letters offering first $25 and then $50 for completion of an interview. In the third interview wave for victims we turned over 70 difficult cases to a licensed private investigator as a last resort. The private investigator used available computer databases to track victims who had moved and provide us with current addresses. He did not confront victims or their acquaintances, and interviews for women he located were conducted by our staff over the phone. Ultimately, this additional tracking methodology added virtually nothing to the interview success rate.

Completion rates Our completion rate with victims was 50% for the first interview, 46% for the second interview, and 50% for the third interview. First interviews with batterers were obtained with 95% of the sample because interviews were obtained when defendants were present at intake in court for the treatment program. Subsequent completion rates were 40% for the second interview and 24% for the third interview. The fact that attrition among victim interviews was substantially lower than among batterers results from the extra lengths (incentives, in-person visits) to which we went in order to obtain the victim interviews.

FINDINGS AND CONCLUSIONS

Our initial analyses showed that men assigned to a group treatment program for batterers were less likely to be the subject of future crime complaints involving the same parties than men assigned to an irrelevant treatment (community service). This difference was most pronounced at six months after group assignment, but held up over a full year (see Table 5).

Subsequent analyses revealed interesting findings about length of treatment. Due to fortuitous circumstances, we wound up splitting our treatment sample into two subsamples distinguished by density of treatment sessions. All batterers randomly assigned to treatment were mandated to attend 39 hours of psycho educational group treatment based upon the Duluth model. However, some batterers received the 39 hours in 26 weekly sessions while others received it in longer biweekly sessions for 8 weeks. The former treatment model maximized time that batterers remained in treatment while the latter reduced the chances that batterers' initial motivation would flag over time.

Our results showed that far more men successfully completed the 8-week group than the 26-week group (see Table 6). Roughly similar proportions of batterers began treatment in the 8-week and 26-week groups. Seventy-seven percent of those assigned to the 8-week groups attended at least one class compared to 71% of those

Table 5: Prevalence of criminal justice incidents involving same victim and perpetrator.

	6 months after assignment*	12 months after assignment**
Batterer treatment (n=190)	10%	15%
Community service (n=186)	22%	26%

* Chi-square (1)=10.43, p.=.001
**. Chi-Square (1)=7.78, p.=.005

Table 9: Prevalence of incidents reported by victims to research interviewers.

Table 6: Attendance in 8 vs 26 week batterers' group			
	No attendance	**Some attendance**	**Graduated**
26-week format (n=129)	29%	44%	27%
8-week format (n=61)	23%	10%	67%

assigned to the 26-week groups. But graduation rates were dramatically different. Sixty-seven percent of the men assigned to the 8-week groups graduated compared to just 27% of those assigned to the 26-week groups.[2] guessed We conclude that a much larger proportion of those assigned to treatment were exposed to the full treatment in the 8-week groups than in the 26-week groups.

We expected, therefore, that men assigned to the 8-week group would have a lower rate of recidivism than men assigned to the 26-week group. However, only the 26-week group was statistically different from the control group on future crime complaints at both 6 and 12 months post-sentence: The 8-week group and the control group were indistinguishable (see Table 7). Victim reports of violence to research interviewers showed a similar pattern, but differences between treatment conditions did not approach statistical significance (see Table 8).

The three-group comparisons also were run using multivariate models, and the results are presented in Appendix A. In the multivariate models, treatment effects were assessed after controlling for the effects of defendant age, ethnicity, marital status, employment status, and arrest history. Although introducing control variables is not, strictly speaking, necessary in analyzing data from experiments, doing so increases the precision of statistical tests (Patel, 1996; Armitrage, 1996). The results of two multivariate models using the number of new

---

[2] Chi-square (1)= 27.72, $p < .001$.

Table 7. Prevalence of criminal justice incidents involving same victim and perpetrator.

	6 months after assignment*	12 months after assignment**
26-week batterer treatment (n=129)	7%	10%
8-week batterer treatment (n=61)	15%	25%
Control (community service) (n=186)	22%	26%

* Chi-square (2)=12.35, p.003
**. Chi-Square (2)=13.13, p.001

Table 8. Prevalence of incidents reported by victims to research interviewers.

	6 months after assignment*	12 months after assignment**
26-week batterer treatment	23% (n=52)	14% (n=66)
8-week batterer treatment	19% (n=26)	18% (n=33)
Control (community service)	21% (n=93)	22% (n=90)

*Chi-Square (2)= 0.15, p=.926
**Chi-Square (2)=1.86, p=.394

incidents reported to criminal justice authorities and the number of new incidents reported by victims to research interviewers support the conclusions in the paragraph above. In addition, an analysis of time to failure using criminal justice data also shows a significant effect of the 26-week treatment.

Finally, we examined measures of the cognitive change in batterers, including conflict resolution skills, beliefs about domestic violence, and locus of control Means and standard deviations for each of the three tests at each of the two time points are presented in Table 9. For each scale, means across the three treatment groups are remarkably similar, and none of the tests shown in Table 9 come close to statistical significance. We have, therefore, no basis for claiming that treatment changed batterers' attitudes or ways of dealing with conflict. But we note that serious limitations in the scales and in our data do not permit an adequate test of this hypothesis. (For a discussion of limitations, the reader is referred to the full report.)

*****

Batterer intervention can be looked upon in one of two ways. It may be a learning process in which attitudes and behavior are modified in a relatively permanent way, Or it may be that batterer intervention simply suppresses violent behavior for the duration of treatment, but no permanent changes are effected. Our results do not support the model of treatment as a change process: If that were true, then the men in the 8-week group (who were finished with

Table 9. Means and Standards deviations for psychosocial outcomes*.

		Conflict Resolution Skills	Attitudes Toward Spouse Abuse	Internal/External Locus of Control
6-month survey	**Control (n=69)**	18.1 (6.3)	25.2 (5.5)	3.5 (2.0)
	8-week (n=27)	19.6 (6.1)	25.2 (6.5)	2.9 (2.4)
	26-week (n=53)	18.0 (5.7)	25.2 (5.1)	3.2 (2.1)
		F(2,116)=0.57	*F(2,146)=0.00*	*F(2,146)=0.41*
12-month survey	**Control (n=37)**	19.3 (6.2)	24.4 (4.1)	3.5 (2.0)
	8-week (n=18)	19.1 (6.0)	25.1 (4.8)	3.1 (2.5)
	26-week (n=33)	19.9 (5.9)	25.9 (4.6)	3.1 (2.1)
		F(2,62)=0.91	*F(2,85)=0.35*	*F(2,85)=0.51*

*Numbers on parentheses are standard deviations.

treatment long before the follow-up period was up) ought to have been as non-violent as their 26-week counterparts (who were in treatment for most of the follow-up period). Yet that is not what our results showed. Also, we did not find evidence that treatment altered attitudes toward spouse abuse, further suggesting that there was no basis for permanent changes. (However, the reader is again advised of serious limitations in the cognitive change scales and data.)

Our results, then support the suppression model of batterer intervention. But they are only suggestive since the study was not designed to test the validity of various models of the treatment process. Moreover, they are at odds with other studies which have not tended to find a difference in recidivism according to length of treatment (Edelson and Syers, 1990; Gondolf, 1997a). Many current batterer programs are going to longer treatment models, but there is also substantial pressure from the defense bar and economics to keep time in treatment to a minimum. Thus, the question of whether treatment works only as long as men attend groups is key to intelligent policy formulation.

How do our results fit into the literature on batterer treatment? If we concentrate only on the four quasi- and two true experiments (including ours), then we note that five of the six (Harrell, et. al. is the lone exception) reported results in the expected direction and all reported statistical significance on at least one outcome measure.

Taken together, these studies provide a case for rejecting the

Taken together, these studies provide a case for rejecting the null hypothesis that treatment has no effect on violent behavior toward spouses. However, the number of useful studies is small and more well-designed studies are warranted before coming to firm conclusions.

Our study provides a good illustration of the difficulties that can be encountered implementing a true experimental design. We had to make substantial concessions to court officials in order to gain their cooperation. Judges were allowed to override assignments to the control group in exceptional cases. This produced a high rate of judicial overrides of cases assigned to the control group. As we showed in the last chapter, the effect of including the override cases in the control group was to make the tests of treatment effects more conservative. (Yet, we still found large treatment effects.) Also, we had to offer a treatment alternative that was more palatable to the defense than the lengthy and costly version that we started with. This proved to be a fortuitous change, however, since we found substantial differences in outcomes between men assigned to the 8-week and 26-week groups. We agree with the opinion of Fagan (1996) and most serious researchers, however, that the benefits of random assignment outweigh the potential difficulties.

APPENDIX A

RESULTS OF MULTIVARIATE ANALYSES

TABLE A-1

Poisson Regression of Annual Rate of Any Officially Recorded Offense

Model Parameters	Model 1 b	Model 1 s.e.	Model 1 Exp(B)	Model 2 b	Model 2 s.e.	Model 2 Exp(B)	Model 3 b	Model 3 s.e.	Model 3 Exp(B)	Model 3 b	Model 3 s.e.	Model 3 Exp(B)
ATV												
Short	-0.24	0.30	0.8	-0.24	0.29	0.8	-0.28	0.46	0.8	0.02	0.35	1.0
Long	-0.58	0.24	0.6 *	-0.57	0.24	0.6 *	-0.30	0.34	0.7	-0.90	0.36	0.4 *
Age				0.00	0.01	1.0	0.00	0.01	1.0	0.00	0.01	1.0
Ethnicity(African-American)												
Hispanic				-0.29	0.25	0.7	-0.28	0.25	0.8	-0.28	0.25	0.8
West Indian/Caribbean				-0.47	0.30	0.6	-0.47	0.30	0.6	-0.45	0.30	0.6
Other Race				-0.33	0.32	0.7	-0.32	0.32	0.7	-0.31	0.32	0.7
Married				0.19	0.20	1.2	0.22	0.20	1.2	0.14	0.26	1.1
Employed				-0.24	0.21	0.8	-0.12	0.26	0.9	-0.28	0.21	0.8
Prior Arrest				0.35	0.20	1.4	0.36	0.20	1.4	0.38	0.20	1.5
ATV * Employment												
Short * Employment							0.07	0.58	1.1			
Long * Employment							-0.52	0.49	0.6			
ATV * Married												
Short * Married										-0.65	0.60	0.5
Long * Married										0.66	0.49	1.9
Intercept	-1.10	0.13	*	-1.08	0.43	*	-1.17	0.44	**	-1.03	0.43	*
Model Fit												
Log likelihood	-241.71			-236.52			-235.88			-234.52		
Restricted Log likelihood	-244.89			-244.89			-244.89			-244.89		
Chi-square	6.36			16.74			18.02			20.75		
P-value	0.04			0.05			0.08			0.04		

TABLE A-2

Negative Binomial Regression of the Past Two Month Frequency of Victimization @ Six Month Survey

	Model 1			Model 2			Model 3			Model 3		
Model Parameters	b	s.e.	Exp(B)	b.	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)
ATV												
Short	-1.53	1.34	0.2	-1.12	1.43	0.3	0.49	2.72	1.6	-2.93	2.16	0.1
Long	-0.88	0.90	0.4	-1.02	0.91	0.4	-0.74	1.12	0.5	-1.05	1.28	0.3
Age				0.04	0.08	1.0	0.06	0.08	1.1	0.05	0.08	1.0
Ethnicity (African-American)												
Hispanic				1.38	1.11	4.0	1.68	1.25	5.4	1.36	1.09	3.9
West Indian/Caribbean				0.32	1.25	1.4	0.11	1.29	1.1	0.13	1.45	1.1
Other Race				0.96	1.84	2.6	1.04	1.96	2.8	0.66	1.96	1.9
Married				-1.39	1.24	0.2	-1.52	1.27	0.2	-1.74	1.45	0.2
Employed				-0.44	1.02	0.6	0.23	1.312	1.3	-0.36	1.04	0.7
Prior Arrest				1.11	1.14	3.0	0.82	1.14	2.3	1.24	1.12	3.5
ATV * Employment												
Short * Employment							-2.63	3.39	0.1			
Long * Employment							-0.68	1.73	0.5			
ATV * Married												
Short * Married										2.58	2.96	13.1
Long * Married										-0.02	1.92	1.0
Intercept	-4.79	7.63		-7.95	6.94		-10.75	8.94		-7.57	6.81	
Selection Bias ratio	6.45	9.07		8.15	8.38		10.34	10.31		7.48	8.18	
Scalar	11.04	3.30	***	9.0	2.46	***	8.81	2.47	***	8.72	2.40	***
Model Fit												
Log likelihood	-199.2177			-193.87			-192.77			-192.77		
Restricted Log likelihood	-545.7404			-474.46			-473.13			-468.86		
Chi-square	694			561.18			560.41			552.18		
P-value	0.00			0.00			0.00			0.00		

TABLE A-3

Negative Binomial Regression of the Past Two Month Frequency of Victimization @ Twelve Month Survey

-	Model 1			Model 2			Model 3			Model 3		
Model Parameters	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)
ATV												
Short	-0.94	1.01	0.4	-0.79	1.18	0.5	-2.16	2.29	0.1	-2.10	1.66	0.1
Long	-1.29	0.81	0.3	-1.57	1.07	0.2	-1.70	1.47	0.2	-0.95	1.12	0.4
Age				0.02	0.05	1.0	0.01	0.05	1.0	0.01	0.05	1.0
Ethnicity(African-American)												1.0
Hispanic				-0.85	1.06	0.4	-0.57	1.29	0.6	-0.51	1.21	0.6
West Indian/Caribbean				0.34	1.18	1.4	0.43	1.44	1.5	0.51	1.36	1.7
Other Race				0.10	1.69	1.1	0.00	1.85	1.0	-0.02	1.61	1.0
Married				-0.86	1.30	0.4	-0.98	1.28	0.4	-0.51	1.17	0.6
Employed				-0.80	1.18	0.4	-1.18	1.41	0.3	-1.00	1.12	0.4
Prior Arrest				-1.03	0.92	0.4	-0.83	0.96	0.4	-0.90	0.93	0.4
ATV * Employment												1.0
Short * Employment							2.06	2.90	7.8			1.0
Long * Employment							0.20	2.11	1.2			1.0
ATV * Married												1.0
Short * Married										1.90	2.47	6.7
Long * Married										-3.52	2.82	0.0
Intercept	1.62	7.93		4.41	9.14		3.56	9.60		5.49	9.21	
Selection Bias ratio	-1.02	10.12		-3.98	11.51		-2.27	12.12		-5.14	11.49	
Scalar	13.92	3.36	***	11.97	3.34	***	11.65	3.36	***	10.35	2.98	***
Model Fit												
Log likelihood	-191.03			-187.30			-186.44			-182.28		
Restricted Log likelihood	-617.99			-551.84			-546.37			-529.49		
Chi-square	853.92			729.09			719.85			694.43		
P-value	0.00			0.00			0.00			0.00		

TABLE A-4

Cox Regression Model of Time-to-first New Officially Recorded Offenses Against Same Victim

	Model 1			Model 2			Model 3			Model 4			Model 5		
Model Parameters	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)	b	s.e.	Exp(B)
ATV															
Short	-0.21	0.29	0.8	-0.52	0.64	0.6	-0.15	0.30	0.9	-0.16	0.47	0.9	0.108	0.36	1.1
Long	-0.72	0.26	0.5 **	-1.36	0.63	0.3 *	-0.74	0.26	0.5 **	-0.75	0.39	0.5 *	-0.96	0.36	0.4 **
ATV															
Short * Time				0.00	0.00	1.0									
Long * Time				0.00	0.00	1.0									
Age							0.01	0.01	1.0	0.01	0.01	1.0	0.01	0.01	1.0
Ethnicity(African-American)															
Hispanic							-0.26	0.26	0.8	-0.26	0.26	0.8	-0.27	0.26	0.8
West Indian/Caribbean							-0.50	0.31	0.6	-0.50	0.31	0.6	-0.50	0.31	0.6
Other Race							-0.76	0.39	0.5 **	-0.76	0.39	0.5 *	-0.74	0.39	0.5 *
Married							0.09	0.22	1.1	0.09	0.22	1.1	0.05	0.26	1.0
Employed							-0.28	0.22	0.8	-0.27	0.26	0.8	-0.32	0.22	0.7
Prior Arrest							0.53	0.22	1.7 **	0.53	0.22	1.7 *	0.55	0.22	1.7 *
ATV * Employment															
Short * Employment										0.01	0.60	1.0			
Long * Employment										0.03	0.53	1.0			
ATV * Married															
Short * Married													-0.65	0.62	0.5
Long * Married													0.50	0.53	1.6

Model Fit	Model 1	Model 2	Model 3	Model 4	Model 5
Log likelihood	1035.47	1035.47	1035.47	1035.47	1035.47
Restricted Log likelihood	1027.11	1025.58	1010.80	1010.79	1008.15
Chi-square	7.90	9.15	24.40	24.80	7.903
P-value	0.02	0.05	0.00	0.01	0.02

REFERENCES

Adams, D. (1988). Counseling men who batter: A profeminist analysis of five treatment models. In M. Bograd & K. Yllo (Eds.), *Feminist perspectives on wife abuse* (pp. 177-198). Beverly Hills, CA: Sage.

Armitage, P. (1996). The design and analysis of clinical trials. In S. Ghosh & C.R. Rao (Eds.) *Handbook of statistics, vol. 13: Design and analysis of experiments.* North-Holland.

Baker, S. & Sadd, S. (1979). *Court employment project final report.* New York: Vera Institute.

Bem, D.J. & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomolous process of information transfer. *Psychological Bulletin, 115,* 4-18.

Berk, R. A. 1983. An introduction to sample selection bias in sociological data. *American Sociological Review* 48(3, June):386-98. June.

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

Brannen, S.J. & Rubin, A. (1996). Comparing the effectiveness of gender-specific and couples groups in a court-mandated spouse abuse treatment program. *Research on Social Work Practice, 6,* 405-424.

Breen, R. 1996. *Regression models: censored, sample-selected, or truncated data.* Sage University Papers Series: Quantitative application in the social science. Thousand Oaks: CA: Sage Publiation.

Buzawa, E., & Buzawa, C. (1996) *Domestic violence: The criminal justice response* (2nd edition). Newbury Park: Sage Publications.

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

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1,* 98-101.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crowell, N., & Burgess, A. W. (Eds.). (1996). *Understanding violence against women.* Washington, DC: National Academy Press.

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

Davis, R.C. & Taylor, B.G. (In press). Does batterer treatment reduce violence? A synthesis of the

Gottfredson, M. R., Gottfredson, D. M. 1988. *Decision making in criminal justice: toward the rational exercise of discretion.* Ed. J. Feinber, T. Hirschi, B. Sales, D. Walker. Law, Society and Policy. New York: Plenum Press.

Gottman, J. M., Jacobson, N. S., Rushe, R. H., Shortt, J. W., Babcock, J., La Taillade, J. J., & Waltz, J. (1995), The relationship between heart rate reactivity, emotionally aggressive behavior, and general violence in batterers. *Journal of Family Psychology, 9*(3), 227-248.

Grusznski, R. J., & Carillo, T. P. (1988). Who completes batterer's treatment groups? An empirical investigation. *Journal of Family Violence, 3*, 141-150.

Hamberger, L. K,. & Hastings, J. E. (1988). Skills training for treatment of spouse abusers: An outcome study. *Journal of Family Violence, 3*, 121-130.

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

Hamberger, L. K., & Hastings, J. E. (1990). Recidivism following spouse abuse abatement counseling: Treatment and program implications. *Violence and Victims, 5*, 157-170.

Hamberger, L. K., & Hastings, J. E. (1993). Court-mandated treatment of men who assault their partners: Issues, controversies, andoutcomes. In N. Z. Hilton (Ed.), *Legal responses to wife assault: Current trends and evaluation*. Newbury Park, CA: Sage.

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

Hardy, M. A. 1993. *Regression with dummy variables*. Sage University Paper series on Quantitative applications in the social sciences. Newbury Park, CA: Sage Publications.

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

Harrell, A. V., Roehl, J. A., & Kapsak, K. A. (1988). *Family violence intervention demonstration programs evaluation, volume II: Case studies.* Report submitted to the Bureau of Justice Assistance. Washington, DC: The Institute of Social Analysis.

Harris, R., Savage, S., Jones, T., & Brooke, W. (1988). A comparison of treatments for abusive men and their partners within afamily-service agency. *Canadian Journal of Community Mental Health, 7*(2), 147-155.

Healey, K., Smith, C., & O'Sullivan, C. (1997). *Batterer intervention: Program approahes and criminal justice strategies.* Report of Abt Associates to the National Institute of Justice, Washington, DC.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1, January):153-61. January.

Holtzworth-Munroe, A., Staurt, G. L. 1994. Typologies of male batterers: three subtypes and the deference among them. *Psychological Bulletin* 116(3):476-97.

Hotaling, G. T., Surgarman, D. B. 1990. A risk maker analysis of assaulted wives. *Journal of Family Violence* 5(1):1-13.

Jacobson, N. S., Gottman, J. M., Shortt, J. W. (1995). The distinction between type 1 and type 2 batterers--further considerations: Reply to Ornduff et al. (1995), Margolin et al. (1995), and Walker (1995). *Journal of Family Psychology, 9*(3), 272-279.

Land, K., C., McCall, P. L., Nagin, D. S. 1996. A comparison of Poisson, negative binomial, and semiparametic mixed regression models. *Sociological Methods and Research* 24(4, May):387-442. May.

Lee, E. T. 1992. *Statistical methods for survival data analysis.* Wiley series in probability and mathematical statistics. Applied probability and statistics. New York, NY: John Wiley & Sons.

Little, R. J. A., Schenker, N. 1995. Missing data. In *Handbook of Statistical Modeling for the Social and Behavior Science*, ed. G. Arminger, C. C. Clogg, M. E. Sobel, pp. 39-76. New York, NY: Plenum Press.

Maiuro, R.D., Cahn, T.S., Vitaliano, P.P. & Zegree, J.B. (1987, August) Treatment for domestically violent men: Outcome and follow-up data. Paper presented at the meeting of the American Psychological Association, New York.

Martin, S., Sechrest, L., & Redner, R. (Eds.) (1981). *New directions in the rehabilitation of criminal offenders*. Washington, D.C.: National Academy of Sciences Press.

Maxwell, C. D. 1998. The specific deterrent effect of arrest on aggression between intimates and spouses. diss. Newark, New Jersey: Rutgers, the State University of New Jersey.

Palmer, S. E., Brown, R. A., & Barrera, M. E. (1992). Group treatment program for abusive Husbands: Long-term evaluation. *American Journal of Orthrpsychiatry, 62*(2), 276-283.

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

Pate, A., Hamilton, E. E. 1992. Formal and informal deterrents to domestic violence. *American Sociological Review* 57(October):691-97. October.

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

Rosenbaum, A., & O'Leary, K. (1986). The treatment of marital violence. In N. S. Jacobsen & A. S. Gurman (Eds.), *Clinical handbook of marital therapy.* NY: Guilford.

Rosenfeld, B. D. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review, 12,* 05-226.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage.

Sampson, R.J. & Laub, J. (1990). Crime and deviance over the life course: The salience of adult social bonds. *American Sociological Review, 55,* 609-627.

Saunders, D. G. (1996a). Interventions for men who batter: Do we know what works. *Psychotherapy in Practice, 2* (3), 81-93.

Saunders, D. G. (1996b). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and Victims.*

Saunders, D. G., & Azar, S. (1989). Family violence treatment programs: Descriptions and evaluation. In L. Ohlin & M. Tonry (Eds.), *Family violence: Crime and justice, a review of research* (pp. 481-546). Chicago, IL: University of Chicago Press.

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

Sherman, L. W., Smith, D. A., Schmidt, J. D., Rogan, D. P. 1992. Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review* 57(October):680-90. October.

Stark, E., Flitcraft, A. 1988. Violence among intimates: an epidemiological review. In *Handbook of family violence*, ed. V. B. Hasselt, R. L. Morrison, A. S. Bellack, M. Hersen, pp. 293-318. New York, NY: Plenum Press.

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

Toby, J. (1957). Social disorganization and stake in conformity: Complimentary factors in the predatory behavior of hoodlums. *Journal of Criminal Law, Criminology, and Police Science, 48,* 12-17.

Tolman, R. M., & Bennett, L. W. (1990). Quantitative research on men who batter. *Journal of*

Interpersonal Violence, 5 (1), 87-118.

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

Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, *6*, 363-378.

Weinstein, G.S. & Levin, B.L. (1989). Effect of crossover on the statistical power of randomized studies. *Annals of Thoracic Surgery*, *48*, 490-95.

Weisberg, S. 1985. *Applied linear regression*. Wiley series in probability and mathematical statistics. Applied probability and statistics. New York, NY: John Wiley & Sons.

Winship, C., Mare, R. D. 1992. Models for sample selection bias. In *Annual Review of Sociology*, pp. 327-50. Palo Alto, CA: Annual Reviews Inc.