

A meta-analytic review of court-mandated batterer intervention programs: Can courts affect abusers' behavior?

LYNETTE FEDER

*Division of Administration of Justice, Portland State University, P.O. Box 751, Portland,
OR 97207-4015, USA
E-mail: lfeder@pdx.edu*

DAVID B. WILSON

*George Mason University, 10900 University Boulevard, MS 4F4, Manassas, VA 20110-2203, USA
E-mail: dwilsonb@gmu.edu*

Abstract. Court-mandated batterer intervention programs are being implemented throughout the United States to address the problem of domestic violence. Prior reviews of research on the effectiveness of these programs have arrived at conflicting conclusions. This study is a systematic review of the extant research on this topic. Experimental and quasi-experimental studies that used matching or statistical controls were included. The results were mixed. The mean effect for official reports of domestic violence from experimental studies showed modest benefit, whereas the mean effect for victim reported outcomes was zero. Quasi-experimental studies using a no-treatment comparison had inconsistent findings indicating an overall small harmful effect. In contrast, quasi-experimental studies using a treatment dropout design showed a large, positive mean effect on domestic violence outcomes. We discuss the weakness of the latter design and raise concerns regarding official reports. The findings, we believe, raise doubts about the effectiveness of court-mandated batterer intervention programs.

Key words: batterer intervention, domestic violence, intimate partner violence, meta-analysis, recidivism

Domestic violence is defined as assaultive behavior involving adults who are married, cohabitating, or who have an ongoing or prior intimate relationship (Goolkasian 1986). Research indicates just how pervasive this problem is today. Based upon crimes reported to the police in 1998, intimate partner homicides accounted for about 11% of all murders nationwide (Rennison and Welchans 2000). The National Crime Victimization Survey (NCVS) indicated that there were about 1 million violent crimes committed against persons by their current or former spouses or significant others in 1998, with the vast majority (85%) being against female victims (Rennison and Welchans 2000).

These numbers demonstrate the extent of the problem in terms of both the amount and severity of violence that some women face. Additionally, research indicates that women who have been victims of domestic violence are at greater risk of future violence (Hilberman 1980; Hirschel and Hutchinson 1992; Langan and Innes 1986). The cost to society is enormous. Domestic crime accounts for almost 15% of the total crime costs – approximately \$67 billion per year

(Miller et al. 1996). This figure does not include the impact that domestic violence has on the children who live in these homes. Past research has established that violent homes are a risk factor for producing violent adults, thereby continuing the 'cycle of violence' (Brisson 1981; Dutton 1988; Widom 1992).

The above figures speak to the importance of finding programs that can successfully intervene with domestic violence offenders. However, individual studies evaluating court-mandated batterer intervention programs have provided very mixed findings on their effectiveness. This systematic review uses meta-analytic procedures to synthesize the extant empirical evidence on the effects that court-mandated batterer intervention programs (including pre-trial diversion programs) have, over and above the effect of routine legal interventions, on rates of recidivism.

Background to court-mandated batterer intervention programs

Decades of overlooking domestic violence as a social problem has recently been followed by an intense amount of public, private, and professional interest in this subject. One of the earliest responses to family violence was the development and growth of shelters for battered women and their children (Johnson and Kanzler 1993). Soon after their establishment, shelter staff noticed that a large percentage of abused women returned to their abusive partners (Hamberger and Hastings 1993; Jennings 1987; Snyder and Scheer 1981). Even where victims successfully separated, these men typically continued their abusive patterns with a different partner (Farley and Magill 1988; Gondolf 1987). These workers came to believe that the best way to stop domestic violence was to change the behavior of the abuser (Feazell et al. 1984).

The original focus of these programs was a direct reflection of their emergence out of the women's shelter movement. Early programs were unstructured groups working with abusive men through a combination of consciousness-raising and peer self-help provided within a context of feminist theory that spoke of men's need to control women (Adams and McCormick 1982; Johnson and Kanzler 1993). Over the next few years, batterer programs developed independently at various sites across the country. As their numbers grew, the earlier unstructured consciousness-raising groups were replaced by more structured groups using psychoeducational and/or cognitive behavioral techniques (Pirog-Good and Stets-Kealey 1985). Still, this was typically done within a feminist context (Healey et al. 1998; Jennings 1987). Most of the programs encouraged men to confront their sexist beliefs and accept responsibility for their past abuse while teaching them alternative behavioral responses like anger management, assertiveness training, relaxation techniques, and communication skills (Davis and Taylor 1999; Healey and Smith 1998).

The Domestic Abuse Intervention Project, out of Duluth, Minnesota (usually just called the Duluth Model), has emerged as one of the more prevalent and widely cited programs for treating battering men. It uses a feminist psychoeducational approach whereby men are taught that battering is part of a range of male behaviors

used to control women. To stop the battering, men are given alternative methods like time-outs, empathizing, problem-solving, and tension-reducing exercises (Pence 1983). The structured curriculum is usually offered in groups that are from 6 to 32 weeks in duration (Tolman and Edleson 1995). The Duluth Model is the model of choice for many communities with some states mandating that batterer intervention programs adhere to this model (Babcock and Taillade 2000).

Cognitive behavioral approaches have also been widely used on this offender population. Typically, batterers (usually in a group format) are offered specific tools to help them see that their acts of violence are not uncontrollable outbursts but rather predictable behavioral patterns that they can learn to stop (Healey et al. 1998). The focus in these groups is on modifying how batterers think and act by working with them on skills training and anger management techniques (Healey and Smith 1998; Tolman and Edleson 1995). Today, most treatment programs blend together aspects of psychoeducational and cognitive behavioral approaches within a feminist context in an attempt to reach a broader range of clientele (Babcock and Taillade 1999; Tolman and Edleson 1995).

Court-mandated interventions using the couple as the unit of treatment are much less widely utilized and have even been expressly prohibited in 20 states (Healey and Smith 1998). Couple counseling sees the couple as the reason for the problem and, as such, works with both members to improve communication and conflict resolution skills. This method has been criticized as blaming victims as well as potentially placing them at greater risk should they honestly express their complaints to the batterer (Babcock and Taillade 1999; Healey et al. 1998).

In 1980, California became the first state to mandate treatment for men convicted of domestic violence (Johnson and Kanzler 1993; Sonkin 1988). The greatest growth in these batterer intervention programs occurred in the late 1980s due to the rise in pro-arrest laws occurring throughout the nation (Hotaling and Sugarman 1986; Johnson and Kanzler 1993). With increasing numbers of jurisdictions presuming or mandating arrest for misdemeanor domestic violence (Dutton and McGregor 1991; Feder 1997), pressure was placed on the courts to deal with these offenders (Ford and Regoli 1993; Pence 1983). At the same time, this population was proving difficult to work with as evidenced by high rates of attrition from these treatment programs (Pirog-Good and Stets-Kealey 1985; Roberts 1982). Having the court mandate treatment was therefore viewed as one method to ensure greater compliance when treating this population (Dutton 1984; Hamberger and Hastings, 1989). Judges also saw this intervention as providing an alternative to prison (important during this period of extensive overcrowding) while simultaneously holding out the hope of breaking the cycle of violence and, in that way, truly helping victims of domestic violence.

Evaluations of court-mandated domestic violence programs

Soon after court-mandated programs began appearing, studies evaluating their effectiveness appeared. In this first wave of evaluation research, the results indicated high rates of success in reducing the frequency and/or severity of

subsequent violence amongst this offender population. However, a number of researchers noted that these findings probably reflected the methodological shortcomings of the research rather than the programs' actual effectiveness in reducing violence (Ford and Regoli 1993; Gondolf 1987). These deficiencies included small sample sizes, failure to study the total population to be evaluated (as opposed to only those who completed the program), lack of appropriate comparison groups, inadequate or variable specification of the primary outcome measures and use of unreliable measures or questionable sources of data to measure treatment outcome (Hamberger and Hastings 1993; Palmer et al. 1992; Tolman and Bennett 1990).

Since then, more rigorous research has been conducted. Unlike the earlier studies, these studies produced mixed results regarding the effectiveness of mandated batterer intervention programs in reducing violence. Over the last 15 years there have been several published reviews of the growing body of research on domestic violence interventions (Babcock et al. 2004; Cromwell and Burgess 1996; Dutton 1988; Eisikovits and Edleson 1989; Hamberger and Hastings 1993; Rosenfeld 1992; Saunders 1996; Tolman and Edleson 1995). As with the individual studies, these reviews offer mixed conclusions regarding the effectiveness of court-mandated batterer intervention programs.

Hamberger and Hastings (1993), after reviewing 28 separate studies, concluded that little is known about the short and long-term effects of these programs. Similarly, Rosenfeld concluded that "the incremental benefit of court-ordered treatment over the deterrent effects of traditional criminal justice system remedies is unclear" (Rosenfeld 1992: 205). More recently, Davis and Taylor (1999) came to a very different conclusion. Computing an average effect size across five experimental or quasi-experimental studies, they concluded that "there is fairly consistent evidence that treatment works and that the effect of treatment is substantial" (Davis and Taylor 1999: 69). Finally, Babcock et al. (2004) conducted a meta-analysis of batterer intervention programs and concluded that, "the effect size due to group battering intervention on recidivism of domestic violence is in the 'small' range" (p. 1043). "To a clinician, this means that a woman is 5% less likely to be re-assaulted by a man who was arrested, sanctioned, and went to a batterers' program than by a man who was simply arrested and sanctioned" (Babcock et al. 2004: 1004).

In 1984, the Attorney General's Task Force on Family Violence recommended court-mandated treatment as an addition to legal alternatives (U.S. Attorney General's Task Force on Family Violence 1984). Yet 20 years later, the field remains uncertain about whether these programs are more effective in reducing future violence than legal interventions alone. The National Academy of Sciences has noted that "the urgency and magnitude of the problem of family violence have caused policy makers, service providers, and advocates to take action in the absence of scientific knowledge that could inform policy and practice" (Chalk and King 1998: 2).

This study attempts to answer this call by conducting a meta-analysis using the most rigorous research on court-mandated batterer intervention programs. Recent research indicates an inverse relationship between design rigor and likelihood of

finding program effectiveness (Feder and Forde 2000; Weisburd et al. 2001). Therefore, we sought to include only the most rigorous research. Like Babcock and her associates, this included experimental designs. Unlike Babcock, we did not include all quasi-experimental studies but instead limited inclusion to those which established pre-treatment equivalence between groups, either via a matched groups design or statistical controls. We also excluded studies that compared one treatment type to another, unless it also included a no-treatment control group. Additionally, in order to be included, the study had to follow offenders for six months post-intervention and use one or more objective measures of repeat violence (i.e., official or victim reports of his continued abuse). The meta-analysis seeks to examine the effect that these court-mandated batterer interventions have on this population's recidivism rate above and beyond what would have been expected through routine legal interventions.

Method

Criteria for inclusion of studies in the review

We sought to assess the effects of post-arrest mandated interventions (including pre-trial diversion programs) in reducing domestic violence offenders' future likelihood of re-assaulting through a synthesis of the extant empirical literature. To be included in this synthesis, a study had to meet the following criteria. First, the study used an experimental or rigorous quasi-experimental design. Experimental designs were defined as those using random assignment to treatment and control group(s). Rigorous quasi-experimental designs were operationalized as those establishing pre-intervention equivalence between the experimental and control group(s) through the use of multivariate statistical methods or a matched subject research design. For both experimental and quasi-experimental designs, no-treatment control groups did not exclude routine treatment by the criminal justice system. That is, no-treatment could include routine legal interventions such as probation, short jail stay, etc., though it would exclude referral to counseling or alternative programs designed specifically to reduce domestic violence (beyond any deterrent effect of jail or probation).

Second, the intervention must have involved a post-arrest court-mandated intervention that, in part or exclusively, was aimed at the batterer and had as its goal decreasing the batterers' future likelihood of re-assaulting that or other partners. As so defined, pre-trial diversion programs could be included in the study as well. Third, only studies using adult participants of heterosexual intimate domestic violence, whether presently or formerly married, separated, divorced, cohabiting or dating were included in the meta-analysis. Fourth, an outcome measure of repeat domestic violence must have been obtained at least six months post-treatment. The decision to follow offenders for a period post-treatment was based on Dunford's findings that evaluation studies collecting outcome data at the end of treatment were more likely to find effectiveness than those measuring

outcomes for some period post-treatment (Dunford 2000). This suggests that evaluations that are based solely on end-of-treatment assessments should be viewed cautiously. Additionally, the study must have included at least one outcome measure on repeat violence to that or other victims that used something more than offenders' self-reported repeat violence. As such, studies could include victim reports on continued abuse or official measures of recidivism including arrest, charges or convictions. Fifth, the studies need to have been conducted in 1986 or later. Finally, the study needed to have reported sufficient data to permit computation of an effect size.

Search strategy for identification of relevant studies

Our goal was to identify and include all studies conducted in the United States or elsewhere from 1986 through January 2003 that met our inclusion criteria. Toward this aim, we searched computerized databases and websites (listed below), bibliographies of published reviews of related literature and scrutinized annotated bibliographies of related literature. We conducted searches of the following databases and websites: ERIC, PsycINFO, MEDLINE, Sociological Abstracts, Social Science Index, Social Work Abstracts, Criminal Justice Abstracts, Social Science Citation Index, Lexis Nexis Legal, Lexis Nexis Medical, Dissertation Abstracts International, GPO Monthly Catalog (MOCAT), National Criminal Justice Research Service, Social, Psychological, Criminological and Educational Trials Register (C2-SPECTR), and the PsiTri database of randomized and controlled trials in mental health.

We used 25 keywords in three clusters to search for all experimental and quasi-experimental studies conducted on the effectiveness of court-mandated interventions for domestic violence offenders. Whenever appropriate a 'wildcard' was used so as to search for the root of the word allowing for other possible derivations. (So, for instance, the term 'eval' was used to pick up evaluation, evaluate, evaluating, etc.) Cluster One related to the subject matter. Cluster Two sought to find citations using program keywords. Finally, Cluster Three used keywords related to outcomes. Terms within a cluster were connected with the Boolean 'or' (i.e., an abstract with any one of the terms would be selected) and the clusters were then connected with the Boolean 'and' (i.e., an abstract with at least one of the terms in each cluster would be selected). To make the resulting list more manageable, the search was restricted to titles and abstracts. If the title or abstract looked promising, the entire study was pulled and reviewed. The keywords within each cluster were: (Cluster One) anger management, batter(er/s), domestic assault, domestic violence, family violence, spous(e/al) abuse, physical abuse, Minneapolis Model, Duluth, or intimate partner violence; (Cluster Two) defer(ral/ring/rred), program(s), treatment(s), intervention(s), diversion(ary), or prosecu(te/tion/torial); and (Cluster Three) effect(s/ive/iveness), research(es), outcome(s), eval(uation/uations/ating), experiment(al), quasi(-experimental), random(ly), compar(ison/ing), or match(ed/es/ing). We examined the bibliographies of the following reviews:

Babcock and Taillade (1999), Babcock et al. (2004), Chalk and King (1998), and Davis and Taylor (1999).

The graduate research assistant and the first author reviewed the titles and abstracted those that were identified through the search process. Studies that appeared likely to be eligible were retrieved in their entirety. Where disagreements occurred, the second author was consulted and differences were resolved. The graduate assistant and first author were also responsible for reviewing the full text of all studies retrieved in their entirety to determine final eligibility for the meta-analysis. Again, where there were disagreements or uncertainties regarding the inclusion of a study, the second author's opinion was sought to resolve the eligibility decision.

The above process identified 11,872 titles and abstracts. (Note: These numbers included duplicates.) Fifty-seven studies were retrieved in their entirety for further scrutiny. Fifteen studies representing 10 distinct experimental (four) or quasi-experimental (six) studies were deemed eligible for the meta-analysis (designated with an asterisk in the reference list).

Coding and data management

Studies determined eligible for inclusion into the systematic review were coded for all relevant data. A four-part coding instrument was used to extract the information (available from the first author). Both authors coded each study and all differences in coding were resolved through negotiation. The coding protocol captured information regarding the nature of the intervention, participant sample, research methods, and outcome results.

To avoid the 'double counting' of findings, two strategies were employed. First, multiple publications of an evaluation were treated as a single study in the synthesis. Second, multiple findings from a single study were categorized by outcome construct and only a single effect per construct was used in any analysis. A decision rule for determining which effect to use in an analysis if multiple effects were available was developed and is explained below in the discussion of results.

Statistical procedures and conventions

This systematic review used standard meta-analytic methods. More specifically, the odds ratio was used as the effect size for dichotomous outcomes, such as official measures of re-arrest, and the standardized mean difference (d) was used for continuous type measures, such as the Conflict Tactics Scales. For ease of presentation, however, the odds ratios were transformed into standardized mean difference type effect sizes. This was done using the methods developed by Hasselblad and Hedges (1995) and involved rescaling the logged odds ratio by a constant. As such, it had no effect on the statistical analyses other than to rescale the values such that they are comparable to the standardized mean difference type effect sizes.

Effects representing unique constructs were analyzed separately. The mean effect size across studies for any given construct was determined by weighting by the inverse variance of the effect size, that is, using the inverse variance weight method. Both fixed and random effects mean effect sizes were computed. The fixed effects model results will only be reported along with the random effects model results if the two models produce substantively different findings (substantially different mean estimate). This is based on an a priori assumption about the distribution of effects across studies that is consistent with the random effects model. These analyses and graphs were performed in Stata with macros written by David B. Wilson that are publicly available <http://mason.gmu.edu/~dwilsonb/ma.html>.

The unit of analysis was the treatment-comparison contrast. Two of the included studies had multiple treatment conditions compared to a single control group (Dunford 2000; Davis et al. 2000). Additionally, Gordon and Moriarty (2003) compared all those mandated to treatment to similar offenders not so mandated and also compared those completing treatment to those who dropped out of treatment. Both of these treatment-comparison contrasts are represented below, although never in the same analysis.

Results

Description of studies

Four experimental studies and six quasi-experimental studies were identified as meeting the eligibility criteria. The basic research design (i.e., randomized, quasi-experimental with a no treatment comparison group, quasi-experimental with treatment dropouts as the comparison group) and treatment type, number of treatment sessions and weeks, nature of the comparison group, and sample description are reported in Table 1.

All 10 studies were conducted in North America. Nine of these studies were published in peer-reviewed journals, although technical reports were also available for four studies (see reference list). When there was conflicting information between the two sources, data from the non-published technical report was used in the coding of the meta-analysis because that typically provided more detailed information.

All 10 studies evaluated a psychoeducational or cognitive behavioral approach, or some mix of the two approaches targeted at the batterer and delivered in all-male group settings. One study (Dunford 2000) also tested two additional intervention types: a cognitive behavioral group targeted at the male batterer but conducted in conjoint groups as well as a no-program but rigorously monitored intervention. In all but two of the studies (Chen et al. 1989; Dunford 2000) it was noted that the program intervention was accompanied by probation, although in one of these studies (Chen et al. 1989) it seems likely that that was the case as well.

Table 1. Description of studies by author and design type.

<i>Author by design type</i>	<i>Treatment type</i>	<i>Treatment sessions/weeks</i>	<i>Comparison type</i>	<i>Sample type</i>
Randomized				
Davis et al. – 8 week program	Psychoeducational	16/8	Probation and 40 hours community service	Convicted batterers – judge, prosecutor and defense must agree to treatment
Davis et al. – 26 week program	Psychoeducational	26/26	Same as above	Same as above
Dunford – men's group	Cognitive – behavioral	32/52	No treatment	Navy sample, incident of domestic violence established, referred to program
Dunford – conjoint	Cognitive – behavioral	32/52	No treatment	Same as above
Dunford – rigorous monitoring		12/52	No treatment	Same as above
Feder and Forde	Cognitive – behavioral/ Psychoeducational	26/26	Probation	All convicted batterers
Palmer et al.	Psychoeducational	10/10	Probation	Convicted batterers – unclear how sample drawn
Quasi-experimental – no treatment comparison				
Chen et al.	Cognitive – behavioral/ Psychoeducational	8 sessions	Non-referred convicted batterers	Convicted batterers referred to treatment program – unclear how sample drawn
Gordon and Moriarty – mandated vs. not	Psychoeducational	22/22	Probation	All convicted batterers
Harrell	Cognitive – behavioral	10/10	Probation	All batterers convicted or given prosecution deferred
Syers and Edleson	Psychoeducational	Batterers not mandated to counseling	All batterers having police contact who could be followed for 12 months	
Quasi-experimental – dropouts as comparison				
Dutton	Cognitive – behavioral	16/16	Treatment dropouts, no-shows and rejects	Convicted batterers – unclear how sample drawn

Table 1. Continued.

<i>Author by design type</i>	<i>Treatment type</i>	<i>Treatment sessions/weeks</i>	<i>Comparison type</i>	<i>Sample type</i>
Jones and Gondolf	Cognitive – behavioral	20/20	Treatment dropouts	Batterers in 4 treatment programs 79% court-mandated/ 21% voluntary referrals
Gordon and Moriarty – completers vs. dropouts	Psychoeducational	22/22	Treatment dropouts	All convicted batterers

Distinct treatment-comparison contrasts within an individual study are listed separately.

The treatment length ranged from a minimum of eight 2-hour sessions (Chen et al. 1989) to a maximum of 32 sessions over the course of a year (Dunford 2000). Treatment length information was not provided by Syers and Edleson (1992). Many of the studies indicated the number of sessions and number of weeks but not the length of the treatment sessions.

The nature of the control group also varied from study to study. The Dunford study (2000) was the most unusual with the control group receiving no intervention whatsoever. Several studies (Feder and Forde 2000;¹ Gordon and Moriarty 2003; Harrell 1991; Palmer et al. 1992) had the control group receiving probation only. The Davis et al. (2000) study used a control group who received 40-hour of community service. The Gordon and Moriarty (2003) study included comparisons to both probation only and treatment no shows and drop-outs. Jones and Gondolf (2002) and Dutton (1986) also used treatment drop-out comparison group designs. Dutton (1986) included men who were rejected from treatment as well as the treatment no-shows and dropouts. Treatment no-shows and drop-outs represented 84% of the sample in Dutton's study and as such is considered a treatment drop-out type study for the analyses below. Finally, one study (Syers and Edleson 1992) did not specify what the control group received beyond not being mandated into counseling.

All but one of the 10 studies used a general civilian population of batterers who were facing or had faced court prosecution for domestic violence. The one exception, Dunford (2000), used men living on a Navy base where an incident of domestic violence had been established and the man had been referred to the program. And all but one study (Jones and Gondolf 2002) used a sample of men who were entirely court-mandated into the batterer program. The Jones and Gondolf (2002) study had a sample that was composed of 79% court-mandated and 21% voluntary clients.

In five studies the representativeness of the sample to the general domestic violence offender population was questionable due to conditions used for inclusion into their sample. In one of the experimental studies (Palmer et al. 1992), inclusion

criteria was suspected of being highly restrictive, in that the resulting sample size was small despite the large jurisdiction from which it was pulled and the long time frame implemented for the study. A second experimental study (Davis et al. 2000) used highly restrictive criteria for inclusion in their sample. In that study, all individuals making up the courtroom workgroup, including the batterer, had to agree to this intervention (versus another non-jail alternative). This, as the researchers noted, led to a pool of more highly motivated offenders than is typically found in the generalized batterer population. In the Dunford study (2000), the men were all living on a naval base with their families and therefore may represent a group with higher stake in conformity than is true of other batterer samples. In one of the quasi-experimental studies (Syers and Edleson 1992) only those men who could be followed 6- and 12-month post-initial police visit were included in the study. This restriction makes it less likely that more marginal batterers would be included in their study. Another quasi-experimental study (Jones and Gondolf 2002) excluded data from one of four sites because the men were deemed at higher risk for subsequent re-offending.

Meta-analytic findings

The effect sizes were analyzed separately by outcome type (official reports and victim reports) and by design type (experimental, quasi-experimental with a no-treatment comparison group, and quasi-experimental with treatment dropouts as the comparison group). Table 2 presents the random-effects mean effect size, 95%

Table 2. Random effects mean effect size (d) and related statistics for official and victim reported measures of domestic violence by design type.

Outcome by design type	Mean <i>d</i>	95% CI		<i>k</i> ^a	<i>Q</i>
		Lower <i>d</i>	Upper <i>d</i>		
Official measures					
Experiments (randomized)	0.26*	0.03	0.50	7	8.19
Quasi-experiments (nonrandomized)					
No treatment comparison group	−0.14	−0.44	0.31	4	12.00*
Treatment dropouts as comparison ^b	0.97*	0.12	1.82	3	12.00*
Victim report measures					
Experiments (randomized)	0.01	−0.11	0.13	6	1.84
Quasi-experiments (no treatment comp.) ^c	−0.11	−0.50	0.27	1	
Total	−0.00	−0.12	0.11	7	2.18

* $p \leq 0.05$.

^aNumber of effect sizes.

^bFixed effects mean effect size was lower (mean $d = 0.49$, 95% CI of 0.27–0.71). Although substantially lower in value, this still represents a large effect in this context and leads to the same substantive conclusions.

^cFixed effect.

confidence interval, and homogeneity statistic (Q) for both outcome types and each design type. The results will be discussed separately for each outcome.

Official reports

Official reports were either official complaints made to the police that may or may not have resulted in an arrest, or actual arrests for domestic violence. If multiple follow-up time points were available, the longest was selected. As can be seen in Table 2, the mean effect size for the experimental (randomized) studies across these 7 comparisons was 0.26. This represents a finding of a moderate reduction in re-offending, with a 95% confidence interval of 0.03–0.50 ($z = 2.23$, $p = 0.03$). Figure 1 indicates a general pattern of positive effects on official reports of repeat victimization in these experimental studies. These estimates varied from a near-zero effect (Davis et al., eight-week program) to large positive effects (Palmer et al., 10-week program; Davis et al., 26-week program). The mean represents a small positive reduction in repeat victimization. This effect roughly represents a reduction in recidivism from 20% to 13%. However, given the small number of studies (four), there is substantial uncertainty regarding the precision of this estimate.

There is additional doubt concerning what the results of one of the studies actually indicates. Specifically, though the Brooklyn Experiment was written as indicating modest support for court-mandated treatment's effectiveness (Davis et al. 2000; Taylor et al. 2001), the findings ran counter to expectations. As noted in their study, treatment completion was higher for the 8-week program than the 26-week program. Yet treatment effects were higher for the men assigned to the 26-week program with an effect size near zero for those assigned to the 8-week program. This differential effect suggests that something other than the batterer program accounted for the positive treatment effect. If the batterer program itself was effective, then the group receiving a higher dose (eight-week program) should have had the better outcome. At the time, Feder and her colleagues speculated that these results were more consistent with a conclusion that supervision, and not

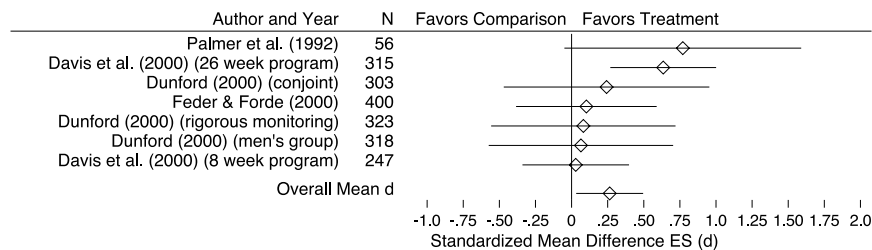


Figure 1. Effect size (d) and 95% confidence interval for official measure from experimental (random) studies.

treatment, resulted in the groups' differences in rates of reassaulting (Feder and Forde 2000; Feder and Dugan 2002). In fact, findings from the Brooklyn Experiment have now come to be viewed by the principal investigator as indicative that additional monitoring and not batterer programs were responsible for differences in recidivism between the three groups (Davis, personal communication). Thus, the strongest empirical evidence for the effectiveness of these programs comes from Palmer et al. (1992), a study with a very small sample size (30 men in the batterer program and 26 in the comparison condition). This small sample size leads to a very unstable estimate of the true treatment effect, as is evident in the rather large confidence interval.

We also noticed in coding the experimental studies that the offender population was restricted in some cases, that is, did not reflect the general domestic violence offender population in two studies – the Palmer and Davis studies (see Table 1). Analyzing the official report effect sizes by this distinction shows a lower non-significant overall mean effect size (0.12, with a 95% CI of -0.21 to 0.44) for the studies using a general domestic violence offender population and a higher mean effect size (0.39, with a 95% CI of 0.10 – 0.67) for the studies with a restricted sample. We are unsure what this finding suggests because the specific restrictions placed on one of these samples (Palmer et al. 1992) were not entirely clear. And, as discussed above, the pattern of results between the 8-week and 26-week programs from the Davis et al. (2000) study are not consistent with the hypothesis that batterer intervention programs are effective.

The quasi-experimental studies represented two fundamentally different design types: designs comparing offenders mandated to treatment to those not mandated and designs comparing treatment completers to treatment drop-outs, no-shows and/or rejects. Because the effect that each design is estimating is different, these two design types were analyzed separately. Table 2 indicates that the mean effect size across the former design (not mandated to treatment comparison) was -0.14 , a small negative effect that is statistically not significantly different from zero. As indicated in Figures 2 and 3, these four credible quasi-experimental studies provide a mixed picture (also evidenced by the significant homogeneity test, Q), with one study observing a moderate positive benefit, one a small positive benefit and two observing a negative effect of a court-mandated treatment relative to a non-mandated group. These estimates statistically adjust for baseline difference although it is unlikely that all of the important differences between the groups

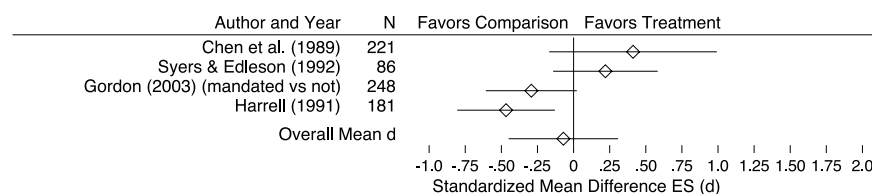


Figure 2. Effect size (d) and 95% confidence interval for official measures from quasi-experimental (nonrandomized) studies with a no treatment comparison group.

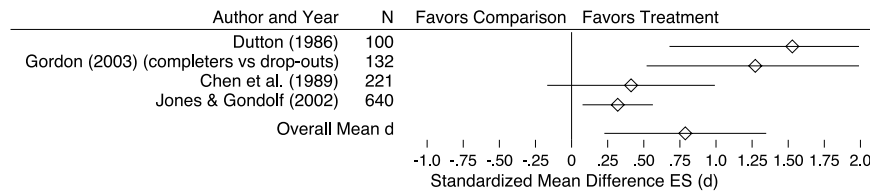


Figure 3. Effect size (d) and 95% confidence interval for official measures from quasi-experimental (nonrandomized) studies with treatment dropouts as comparison.

were taken into account. The composite or mean effect has a plausibility range extending from a small negative to a small positive effect.

The second quasi-experimental design type compared batterers who completed a court-mandated treatment program with those who were mandated and were either rejected from treatment, never showed or dropped out. The three studies with this design consistently found a positive and significant effect. That is, abusers mandated to a domestic violence treatment who complete their program re-offend at a substantially lower rate than offenders who were mandated to these programs who did not complete their treatment. Unfortunately, we cannot attribute this difference solely to the impact of treatment, as treatment attendance is likely to be confounded with other important variables. That is, men who attend and complete their treatment may be more highly motivated to change or more fearful of further criminal justice involvement than men who do not complete a treatment program that has been judicially mandated. Differences in rates of recidivism may be attributed, then, to differences existing in the groups prior to the intervention. In other words, the relationship that we think we are observing between treatment non-completion and recidivism may be spurious and due to another, unobserved variable.

Victim reported outcomes

A concern with official measures is that they may not accurately reflect the amount and severity of ongoing violence. Research consistently indicates that official reports capture only a small fraction of this abuse (Dutton 1988; Straus 1991; Tjaden and Thoennes 2000). As such, the victim is viewed as the best source for information on the offender's continued abuse. Given that, we turn our attention to the seven estimates we have from these studies on the effect of these programs from victim reports of abuse. Three of the four experimental studies measured the victim's reports of their partner's abusive behavior using either the standardized Conflict Tactics Scale (CTS) or the modified Conflict Tactics Scale (CTS2) (Straus et al. 1996). One of the quasi-experimental studies also measured the victim's report of their partner's abusive behavior using a measure similar to

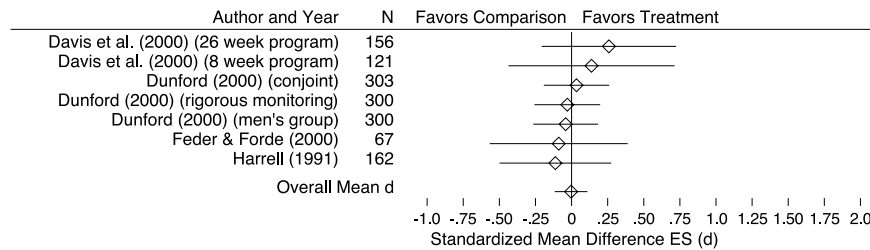


Figure 4. Effect size (d) and 95% confidence interval for victim reported measures from experimental and quasi-experimental studies with a no treatment comparison group.

the CTS. For purposes of analysis, we coded all reported subscales and averaged the multiple effect sizes within each treatment-comparison contrast, with the exception of Harrell (1991), where we selected the outcome based on the largest portion of the sample (necessitated by the way in which the researcher reported the data). Thus, the effect size used in Table 2 and Figure 4 represents the mean effect across subscales of the CTS/CTS2 for the comparison of interest. As shown in Table 2, the mean effect size for victim reports in studies using an experimental design was near zero and was not statistically significant. The effect size for quasi-experimental studies showed a small and negative effect for treatment though this finding also was not statistically significant. The distribution of effects is shown in Figure 4. Three of these effects are positive, four are negative, and none are statistically significant. Thus, the outcome measures based on the female intimate partner's report, and the more credible of the quasi-experimental studies, do not replicate the finding of a small but positive benefit of treatment found in the experimental studies using the official measures of re-offending.

Discussion

This systematic review was based on 10 experimental and quasi-experimental studies. The experimental studies looked at the effect of mandating batterer intervention programs relative to a no-treatment or routine-treatment approach for men facing or convicted of misdemeanor domestic violence charges. Two of the quasi-experimental studies compared men court-mandated to counseling with those not court-mandated (Syers and Edleson 1992; Harrell 1991), two compared court-mandated men who completed treatment to those mandated who did not complete treatment (Dutton 1986; Jones and Gondolf 2002) and one study (Gordon and Moriarty 2003) included both comparisons. All of the evaluated programs used a psychoeducational, feminist oriented and/or cognitive behavioral approach.

The evidence from our meta-analysis is mixed. There is some support for the modest benefits of batterer programs from official reports in the experimental studies, but this effect is smaller if we look only at studies using a general batterer population. Additionally, the effect is absent when victim reported measures are examined. The quasi-experimental studies using a no-treatment comparison also fail to find a positive treatment effect in terms of a reduction in violence when measured with official reports. Finally, quasi-experimental studies using men who were rejected from treatment or who rejected treatment showed a large, positive and significant effect on reducing re-offending. However, we have serious concerns about these latter studies as discussed in detail below.

Our findings differ somewhat from those of Babcock et al. (2004). They concluded, based on their meta-analysis, that these programs have a small but positive effect on abusive behavior. There are several differences between the methods employed in our respective meta-analysis that may account for the differing conclusions. Primarily, Babcock et al. did not separately analyze studies using treatment drop-out designs from other quasi-experimental designs, potentially upwardly biasing the mean effect size for these studies. Both of our results based exclusively on experimental studies are fairly consistent. Babcock et al. reported an effect size of 0.12 when using official reports (fixed effects 95% CI of 0.02–0.22). This is somewhat smaller than our overall mean effect for official reports based on experimental studies but consistent with our estimate from those studies with a representative population. Similarly, Babcock and Steiner indicate a treatment effect of 0.09 (fixed effects 95% CI of –0.02 to 0.21) for victim reported outcomes, slightly higher than our estimate (0.01) but neither estimate is statistically significant.

We have four main concerns regarding the findings. First, we question the generalizability of these findings to general convicted batterer populations. Second, we believe there is a potential bias inherent in using official records to measure continued abuse. Third, the victim reports suffer from low reporting rates in these studies, raising concerns about the validity of these estimates. Finally, we question the validity of the quasi-experimental studies that compare treatment completers to rejects, no-shows and dropouts. Each of these concerns is addressed below.

We judged two studies (Davis et al. 2000; Palmer et al. 1992) as having samples that were restricted in a manner that reduced the representativeness of their findings to a general batterer population. Studies that did not have restrictions limiting who was included in the batterer program probably better represented the ‘typical’ convicted batterer. Our analysis indicates that these latter studies had a lower overall mean effect size for official reports of domestic violence than the studies using a restricted sample. Importantly, the mean effect for the more representative studies was not statistically significant, raising the possibility that the overall positive findings of Figure 1 was in part a function of a restricted (possibly more motivated or perhaps ‘creamed’) sample of batterers. This may indicate that batterer intervention programs work for a selected (presumably more motivated) subset of offenders. The evidence on this issue is weak for two reasons:

(1) we do not actually know the motivation levels of the men in the different studies, and (2) the Davis et al. study had inconsistent results across two similarly motivated groups receiving the same intervention, differing only in the number of weeks over which the program was spread. Thus, we believe that there is insufficient data for any strong conclusion on this issue.

The heavy reliance on official measures in all of these studies is also highly problematic. Official measures are dependent on a victim's willingness to file a complaint or call the police. This raises the possibility that assignment to court-mandated treatment versus a no-treatment control group may differentially affect the victim's willingness to contact criminal justice officials when future abuse occurs. (What Cook and Campbell (1979) refer to as an instrumentation effect.) A victim may not report her partner's abuse for a number of reasons. This includes the possibility that she might prefer to see her partner continue in treatment where she believes it will eventually lead to changes in his abusive behavior rather than take the risk of reporting his continued abuse and see him go to jail. Alternately, a victim may resent the criminal justice system's intrusion into her life in the form of mandating a treatment that she is then responsible to pay for. Most programs require the abuser to pay for the treatment and by extension that means that it is the family that pays for the treatment (Zorza 2003). If the treatment is viewed by a victim as ineffective, it may make her critical and suspicious of the system and less likely to cooperate in the case of reporting future incidences of abuse. We have no empirical evidence that this occurs but the dependence of official reports on the behavior of the victim allows for the possibility that the different rates noted between batterers in the treatment and comparison conditions may reflect a measurement artifact and not a genuine treatment effect. This possibility is strengthened by the different findings obtained in these studies depending upon whether official reports or victim reports are used as the outcome measure.

The high rate of victim attrition in many of these studies is another concern. The victim is usually viewed as the best source for information on the offender's continued abuse. Victim reports of abuse via standardized measures such as the Conflict Tactics Scale are less likely to be affected by the issues raised regarding official reports of continued abuse, provided that the victim is convinced of the confidential nature of her responses. Unfortunately, the percentage of victims responding to follow-up surveys in these studies is low, seriously undermining their utility in establishing the effectiveness (or ineffectiveness) of these programs.

The attrition for victim report for the effect sizes shown in Figure 4 was roughly 30% for the Dunford (2000), roughly 50% for the Davis et al. (2000), roughly 80% for the Feder and Dugan (2002) and 59% for Harrell (1991). High attrition raises the possibility that the victims lost to follow-up in the treatment group may differ in meaningful ways from those lost to follow-up in the control group. Thus, the absence of an effect for the victim report measures may reflect that the programs are truly ineffective or, alternately, that there is a positive or negative effect that is masked by differential attrition.

The problem of high rates of victim attrition becomes critical in light of research indicating that certain victims of domestic violence are more likely to be

lost in the research follow-up than are others. This research strongly suggests that women victims of domestic violence who are more difficult to retain in follow-up research are both more marginal and more likely to be more frequently and severely abused (Sullivan et al. 1996). There is also research that indicates that men who are more marginal are both less likely to obey a court-mandate to treatment and more likely to continue to abuse their partners (Feder and Dugan 2002). If we can assume that more marginal women are more likely to be partnered with more marginal men, then the need for maintaining contact with a high percentage of victims when assessing the effectiveness of these spouse abuse abatement programs becomes even more apparent. This may be important to the extent that some research has indicated that factors associated with the abuser's stake in conformity is associated with the likelihood that an intervention will be successful in reducing subsequent violence (Berk et al. 1992; Sherman 1992). At best, this attrition reduces the generalizability of the findings from victim reported outcomes to a subset of the domestic violence offender population. At worst, there may be differential loss of these marginal women from the treatment and control groups, producing bias in the findings.

Finally, we note the difficulty with using treatment dropouts as a control group, even once statistical controls have been introduced. Two specific problems occur with this type of study design, one with the construct of what is being evaluated or tested and the other with the adequacy of the statistical models in adjusting for initial group differences. First, these studies are trying to estimate the affect of full participation in the batterer intervention program above and beyond the court mandate. In other words, they look to answer the question, "Among men who are court-mandated to batterer intervention, do those who choose to attend and complete this program do better than those who do not?" Although this may be of interest to program providers and developers, it does not address the broader issue of the likely reduction in domestic violence as a function of a policy to mandate such treatment. That is, "whether court-mandated batterer intervention programs reduce offenders' likelihood of re-offending." Addressing the latter question is critical to knowing whether court-mandated domestic violence interventions are beneficial to society.

Second, that these studies produce treatment effect estimates that are large given the population and nature of the problem clearly establishes that men who complete these programs recidivate at a lower rate than men who do not. The question is what to make of this empirical finding. The statistical models employed by these studies attempt to adjust for selection differences between the groups of men. To produce unbiased estimates, however, these models need to fully account for the selection process, that is, the reasons why some men attended treatment and others did not. We do not believe that these equations adequately model the selection process. Potentially important variables, such as motivation for treatment, were not included. The positive treatment effect estimate may indicate that the treatment is effective for motivated offenders, though we cannot conclude this since we do not have comparisons in any of these quasi-experimental studies with motivated offenders who were not mandated and did not receive treatment. Equally

plausible, these findings may simply reflect that the subset of offenders who will complete mandated treatments are less likely to re-offend, with or without the treatment (i.e., these programs may have ‘creamed’ those offenders who are least likely to re-offend regardless of what action is taken).

Conclusion

The findings from this meta-analysis combined with the caveats above raise questions as to the value of these programs. While additional research is needed, the meta-analysis does not offer strong support that court-mandated treatment to misdemeanor domestic violence offenders reduces the likelihood of further reassault.

Research implications

The research implication growing out of this synthesis is that additional experiments need to be conducted to more clearly decipher the effectiveness of court-mandated batterer intervention programs. If we are to test the ability of courts to mandate change, these future experiments must ensure samples of batterers that are representative of the larger convicted batterer population rather than a smaller subset of selected batterers. Additionally, these studies must attend to the importance of maintaining high victim retention, so as to better ascertain any positive or negative effects from this mandated intervention. Finally, additional research is needed to better understand the validity and reliability of official report and victim report measures used in these studies and how they might be affected by treatment assignment.

Policy implications

Intervening in the lives of others is a risky business, particularly when the individuals participating in the social intervention are mandated by a court of law to do so. As such, it is incumbent upon us to ensure that we are not inadvertently making things worse for those we are seeking to help. At this point, the existing evidence cannot ensure that these programs are, in fact, helpful and not harmful.

There is no doubt that “There is a tremendous sense of urgency and alarm in the treatment of domestic violence – and rightly so. After all, protecting the physical and emotional safety of women and their children is the first priority. Consequently, clinicians feel a primary obligation to ‘do something immediately and decisively to halt and prevent violence’” (Jennings 1987: 204). But as the above review has indicated, doing something may not help. As McCord so wisely noted, “Unless social programs are evaluated for potential harm as well as benefit,

safety as well as efficacy, the choice of which social programs to use will remain a dangerous guess" (McCord 2003: 16).

It is clear that we need to be guided by rigorous research in helping us set our course. While better research is needed to determine the effectiveness of court-mandated batterer intervention programs, the results from the meta-analysis do not provide confidence that these programs will be found to be effective. Therefore, it would prove beneficial for the criminal justice system to begin looking at other types of interventions for addressing the problem of domestic violence. However, these interventions must be tied to rigorous evaluations to determine their full impact. In other words, we recommend the use of pilot studies joined to an experimental design, as was suggested almost 20 years ago by Berk et al. (1985), as the preferred path for finding effective programs that can effectively meet the challenge that intimate partner violence presents. Such a course would be especially prudent in these times of limited resources. More than that, victims and taxpayers deserve such evidence-based decision making.

Unfortunately, what we are suggesting is not possible in many jurisdictions today in that their statutes require that, upon conviction for domestic violence, individuals must be mandated into a batterer intervention program, not atypically based upon the Duluth Model (Babcock and Taillade 2000). The end result is that judges, prosecutors and probation officers continue to send batterers to these treatment programs, even if they have grave doubts about their effectiveness. Alternate programs cannot be implemented and tested even as evidence builds indicating that batterer intervention programs, at least as designed and implemented today, may not be effective.

Acknowledgements

We thank Sabrina Austin for her work searching, retrieving, and reviewing for eligibility the studies used in this meta-analysis.

This work was supported in part by funding from the Smith Richardson foundation.

Note

- 1 The first listed author on this review (Lynette Feder) was the primary investigator of the Broward Experiment assessing the effectiveness of batterer intervention programs in South Florida. To best counter the potential conflict of interest, the review was done as transparently as possible. Additionally, the researcher chose to collaborate with a colleague to ensure against any bias. Finally, it was decided before beginning the project that where there were disagreements between the first and second authors, an expert arbitrator would be brought in to resolve any conflicts.

References

- Adams, D. & McCormick, A. (1982). Men unlearning violence: A group approach based on the collective model. In M. Roy (Ed.), *The abusive partner: An analysis of domestic battering* (pp. 170–197). New York: Van Nostrand-Reinhold.
- Babcock, J. C. & Taillade, J. (2000). Evaluating interventions for men who batter. In J. Vincent & E. Jouriles (Eds.), *Domestic violence: Guidelines for research-informed practice* (pp. 37–77). Philadelphia: Jessica Kingsley.
- Babcock, J. C., Green, C. E. & Robie, C. (2004). Does batterers' treatment work? A meta-analytic review of domestic violence treatment. *Clinical Psychology Review* 23(8), 1023–1053.
- Berk, R., Boruch, T., Chambers, F., Rossi, P. & Witte, S. (1985). Social policy experimentation: A position paper. *Evaluation Review* 9(4), 387–429.
- Berk, R., Campbell, A., Klap, R. & Western, B. (1992). The deterrent effect of arrest in incidents of domestic violence: A Bayesian analysis of four field experiments. *American Sociological Review* 57(5), 698–708.
- Brisson, N. (1981). Battering husbands: A survey of abusive men. *Victimology* 6, 338–344.
- Chalk, R. & King, P. (1998). *Violence in families: Assessing prevention and treatment programs*. Washington, DC: National Academy Press.
- *Chen, H., Bersani, C., Myers, S. & Denton, R. (1989). Evaluating the effectiveness of a court sponsored treatment program. *Journal of Family Violence* 4, 309–322.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cromwell, N. & Burgess, A. (1996). *Understanding violence against women*. Washington DC: National Academy Press.
- Davis, R. & Taylor, B. (1999). Does batterer treatment reduce violence? *Women and Criminal Justice* 10, 69–93.
- *Davis, R. C., Taylor, B. G. & Maxwell, C. D. (2000). *Does batterer treatment reduce violence? A randomized experiment in Brooklyn*. Washington, DC: National Institute of Justice.
- *Dunford, F. W. (2000). The San Diego Navy experiment: An assessment of interventions for men who assault their wives. *Journal of Consulting and Clinical Psychology* 68, 468–476.
- *Dutton, D. (1984). Interventions into the problem of wife assault: Therapeutic, policy and research implications. *Canadian Journal of Behavioral Science* 16(4), 281–297.
- Dutton, D. (1986). Wife assaulter's explanations for assault: The neutralization of self-punishment. *Canadian Journal of Behavioral Science* 18(4), 381–390.
- Dutton, D. (1988). Research advances in the study of wife assault: Etiology and prevention. *Law and Mental Health* 4, 161–220.
- Dutton, D. & McGregor, B. (1991). The symbiosis of arrest and treatment for wife assault: The case for combined intervention. In M. Steinman (Ed.), *Woman battering: Policy responses* (pp. 131–154). Cincinnati, OH: Anderson Publishing Company.
- Eisikovits, Z. & Edleson, J. (1989). Intervening with men who batter: A critical review of the literature. *Social Service Review* 63, 384–414.
- Feder, L. (1997). Domestic violence and police response in a pro-arrest jurisdiction. *Women and Criminal Justice* 8(4), 79–98.
- *Feder, L. & Dugan, L. (2002). A test of the efficacy of court mandated counseling for domestic violence offenders: The Broward Experiment. *Justice Quarterly* 19(2), 343–375.

- Feder, L. & Forde, D. (2000). *A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward Experiment* (Final report, Grant NIJ-96-WT-NX-0008). Washington, DC: National Institute of Justice.
- Farley, D. & Magill, J. (1988). An evaluation of a group program for men who batter. *Social Work With Groups* 11(3), 53–65.
- Feazell, C., Mayers, R. & Deschner, J. (1984). Services for men who batter: Implications for programs and policies. *Family Relations* 33, 217–223.
- Ford, D. & Regoli, M. J. (1993). The criminal prosecution of wife assaulters. In Z. Hilton (Ed.), *Legal responses to wife assault: Current trends and evaluation* (pp. 127–164). Newbury Park, CA: Sage.
- Gondolf, E. (1987). Evaluating programs for men who batter: Problems and prospects. *Journal of Family Violence* 2(1), 95–108.
- Goolkasian, G. (1986). *Confronting domestic violence: The role of criminal court judges*. Washington, DC: National Institute of Justice.
- *Gordon, J. A. & Moriarty, L. J. (2003). The effects of domestic violence batterer treatment on domestic violence batterer treatment and domestic violence recidivism. *Criminal Justice and Behavior* 30(1), 118–134.
- Hamberger, L. K. & Hastings, J. (1989). Counseling male spouse abusers: Characteristics of treatment completers and dropouts. *Violence and Victims* 4(1), 275–286.
- Hamberger, L. K. & Hastings, J. (1993). Court-mandated treatment of men who assault their partner. In Z. Hilton (Ed.), *Legal responses to wife assault: Current trends and evaluation* (pp. 188–229). Newbury Park, CA: Sage.
- *Harrell, A. (1991). *Evaluation of court-ordered treatment for domestic violence offenders* (Final report). Washington, DC: National Institute of Justice.
- Hasselblad, V. & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin* 117, 167–178.
- Healey, K. & Smith, C. (1998). *Batterer programs: What criminal justice agencies need to know*. Washington, DC: National Institute of Justice.
- Healey, K., Smith, C. & O'Sullivan, C. (1998). *Batterer intervention: Program approaches and criminal justice strategies*. Washington, DC: Department of Justice.
- Hilberman, E. (1980). Overview: The “wife-beater’s wife” reconsidered. *American Journal of Psychiatry* 137(11), 1336–1347.
- Hirschel, J. D. & Hutchinson, I. (1992). Female spouse abuse and the police response: The Charlotte, North Carolina Experiment. *Journal of Criminal Law and Criminology* 83(1), 73–119.
- Hotaling, G. & Sugarman, D. (1986). An analysis of risk markers in husband to wife violence: The current state of knowledge. *Violence and Victims* 1(2), 101–124.
- Jennings, J. (1987). History and issues in the treatment of battering men: A case for unstructured group therapy. *Journal of Family Violence* 2(3), 193–213.
- Johnson, J. & Kanzler, D. (1993). Treating domestic violence: Evaluating the effectiveness of a domestic violence diversion program. *Studies in Symbolic Interaction* 15, 271–289.
- *Jones, A. S. & Gondolf, E. W. (2002). Assessing the effect of batterer program completion on reassault: An instrumental variables analysis. *Journal of Quantitative Criminology* 18(1), 71–98.
- Langan, P. & Innes, C. (1986). *Preventing domestic violence against women*. Washington, DC: National Institute of Justice.
- McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Science* 587, 16–30.
- Miller, T., Cohen, M. & Wiersema, B. (1996). *Victim costs and consequences: A new look*. Washington, DC: National Institute of Justice.

- *Palmer, S., Brown, R. & Barrera, M. (1992). Group treatment program for abusive husbands: Long-term evaluation. *American Journal of Orthopsychiatry* 62(2), 276–283.
- Pence, E. (1983). The Duluth Domestic Abuse Intervention Project. *Hamline Law Review* 6, 247–275.
- Pirog-Good, M. & Stets-Kealey, J. (1985). Male batterers and battering prevention programs: A national survey. *Response* 8, 8–12.
- Rennison, C. R. & Welchans, S. (2000). *Intimate partner violence*. Washington, DC: National Institute of Justice.
- Roberts, A. (1982). A national survey of services for batterers. In Maria Roy (Ed.), *The abusive partner: An analysis of domestic battering* (pp. 230–243). New York: Van Nostrand-Reinhold.
- Rosenfeld, B. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review* 12, 205–226.
- Saunders, D. (1996). Interventions for men who batter: Do we know what works? *In Session: Psychotherapy in Practice* 2(3), 81–93.
- Sherman, L. (1992). The influence of criminology on criminal law: Evaluating arrests for misdemeanor domestic violence. *Journal of Criminal Law and Criminology* 83, 1–45.
- Snyder, D. & Scheer, N. (1981). Predicting disposition following brief residence at a shelter for battered women. *American Journal of Community Psychology* 9, 559–566.
- Sonkin, D. J. (1988). The male batterer: Clinical and research issues. *Violence and Victims* 3(1), 65–79.
- Straus, M. (1991). Conceptualization and measurement of battering: Implications for public policy. In M. Steinman (Ed.), *Woman battering: Policy responses* (pp. 19–47). Cincinnati, OH: Anderson.
- Straus, M., Hamby, S., Boney-McCoy, S. & Sugarman, D. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues* 17(3), 283–316.
- Sullivan, C., Rumptz, M., Campbell, R., Eby, K. & Davidson, W. (1996). Retaining participants in longitudinal community research: A comprehensive protocol. *Journal of Applied Behavioral Science* 32(3), 262–276.
- *Syers, M. & Edleson, J. (1992). The combined effects of coordinated criminal justice intervention in woman abuse. *Journal of Interpersonal Violence* 7, 490–502.
- Taylor, B., Davis, R. & Maxwell, C. (2001). The effects of a group batterer treatment program: A randomized experiment in Brooklyn. *Justice Quarterly* 18(1), 171–201.
- Tjaden, P. & Thoennes, N. (2000). Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the National Violence Against Women Survey. *Violence Against Women* 6(2), 142–161.
- Tolman, R. & Bennett, L. (1990). A review of quantitative research on men who batter. *Journal of Interpersonal Violence* 5, 87–118.
- Tolman, R. & Edleson, J. (1995). Intervention for men who batter: A review of research. In S. Stith & M. Straus (Eds.), *Understanding partner violence: Prevalence, causes, consequences and solutions* (pp. 262–273). Minneapolis, MN: National Council on Family Relations.
- Weisburd, D., Lum, C. & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science* 578, 50–70.
- Widom, C. S. (1992). *The cycle of violence*. Washington, DC: US Department of Justice.
- Zorza, J. (2003). New research: Broward County Experiment shows no benefit from batterer intervention programs. *Domestic Violence Report* 8, 23–25.

About the authors

Lynette Feder, Ph.D. is an Associate Professor in the Criminal Justice and Criminology division at Portland State University. Her research interests are directed at rigorously evaluating criminal justice interventions to address both policy questions as well as underlying theoretical issues in areas including juvenile delinquency, discretion and discrimination in the criminal justice system, and domestic violence.

David B. Wilson, Ph.D., is an Assistant Professor in the Administration of Justice Program at George Mason University. His research interests are the effectiveness of offender rehabilitation and crime prevention efforts, program evaluation methodology, quantitative methods, and meta-analysis.