



Pergamon

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

CLINICAL  
PSYCHOLOGY  
REVIEW

Clinical Psychology Review 23 (2004) 1023–1053

# Does batterers' treatment work? A meta-analytic review of domestic violence treatment

Julia C. Babcock<sup>a,\*</sup>, Charles E. Green<sup>a</sup>, Chet Robie<sup>b</sup>

<sup>a</sup>*Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204-5022, USA*

<sup>b</sup>*College of Business Administration, Niagara University, Niagara Falls, NY, USA*

Accepted 29 July 2002

## Abstract

This meta-analytic review examines the findings of 22 studies evaluating treatment efficacy for domestically violent males. The outcome literature of controlled quasi-experimental and experimental studies was reviewed to test the relative impact of Duluth model, cognitive-behavioral therapy (CBT), and other types of treatment on subsequent recidivism of violence. Study design and type of treatment were tested as moderators. Treatment design tended to have a small influence on effect size. There were no differences in effect sizes in comparing Duluth model vs. CBT-type interventions. Overall, effects due to treatment were in the small range, meaning that the current interventions have a minimal impact on reducing recidivism beyond the effect of being arrested. Analogies to treatment for other populations are presented for comparison. Implications for policy decisions and future research are discussed.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Duluth model; Cognitive-behavioral therapy; Recidivism; Violence

## 1. Introduction

As an estimated 840,000 women reported assaults at the hands of an intimate in 1996 (Bureau of Justice Statistics, 1998), interventions designed to address this growing public

\* Corresponding author.

*E-mail address:* [jbabcock@uh.edu](mailto:jbabcock@uh.edu) (J.C. Babcock).

health concern have focused on the perpetrators of domestic violence in hopes of deterring further assault. Prior to the 1980s, little attention was paid to domestic violence intervention (Fagan, 1989). Issues of family privacy vs. societal best interest were paramount (Zimring, 1989); domestic violence was sometimes thought best “left behind drawn curtains” (*State v. Oliver*, 1874, cited in Rosenfeld, 1992). Subsequent criminalization of domestic violence dictated whether the crime of domestic violence should entail rehabilitation or incarceration. Since then, spouse abusers have “traditionally fallen under the rehabilitative, rather than the punitive arm of the criminal justice system” (Rosenfeld, 1992, p. 207). In actuality, with the implementation of mandatory arrest policies and court-mandated counseling, batterers’ interventions became a fusion between punishment and rehabilitation.

### *1.1. Current standards of care*

While interventions for batterers are far from standardized, standards of care of battering interventions have been evolving in the United States since the 1990s (see Austin & Dankwort, 1999, for a review). Most states target the perpetrator as solely responsible for the crime and, as such, he shall be held accountable. Most guidelines also require training of group facilitators and experience in domestic violence work, although professional degrees and licensure are generally not required. The recommended duration of intervention ranges from 12 to 52 weeks. Finally, the group intervention model is the format of choice in 90% of mandates, and individual and couples’ therapy is deemed as inappropriate in the majority of the current standards (Austin & Dankwort, 1999). For the most part, state standards have been developed independently of empirical research.

Despite declarations that arrest followed by court-ordered treatment offers “great hope and potential for breaking the destructive cycle of violence” (U.S. Attorney General’s Task Force on Family Violence, 1984, p. 48), there is little empirical evidence that treatment is effective in reducing recidivism of family violence to any meaningful degree. In his review of the earlier studies on marital violence treatment programs, Rosenfeld (1992) concluded that men who are arrested and complete treatment have only slightly lower recidivism rates than men who are arrested but refuse treatment, dropout of treatment, or remain untreated. Some have even argued that treatment programs may put women at increased risk for domestic violence, by contributing to a false sense of security among battered women whose husbands have sought treatment (Holtzworth-Munroe, Beatty, & Anglin, 1995).

Fortunately, in the past decade, several researchers have conducted well-designed studies capable of shedding some light on questions and concerns regarding the efficacy of batterers’ treatment. A small but growing body of methodologically rigorous investigations into the effectiveness of current programs now exists. The purpose of this article is to critically review the treatment outcome research on batterers’ interventions and to conduct a meta-analysis to examine the impact of (1) the treatment type and (2) the study design on the effect size attributable to treatment. Since the current community response to battering is a combination of legal sanctions plus rehabilitation, the goal of this meta-analysis is to examine the effect of the therapeutic intervention, over and above the effect of legal interventions.

A number of studies have summarized the effects of batterers' treatment (Babcock & LaTaillade, 2000; Davis & Taylor, 1999; Hamberger & Hastings, 1993; Levesque & Gelles, 1998; Rosenfeld, 1992; Tolman & Bennett, 1990). After their review of the research literature, Hamberger and Hastings (1993, p. 220) asked the question, "What do we know about the short- and long-term effects of treatment on wife assault?" They conclude "Not much," due to methodological problems of the existing research. In his quantitative review, Rosenfeld (1992) concluded that there are minimal decreases in recidivism rates between treatment completers (36%) and men only receiving legal-system interventions (39%). Rosenfeld stopped short of conducting a meta-analysis, due to the limited number of studies using consistent methodologies available at that time. Davis and Taylor (1999) recently reviewed the empirical batterers' treatment outcome literature and came to quite different conclusions. Although they did not conduct a meta-analysis, they calculated the average effect sizes from five studies. Based on these averages, they estimated the treatment effect size to be approximately  $h=0.41$  (less than 0.50 is considered "small") but nonetheless concluded that "there is fairly consistent evidence that treatment works and that the effect of treatment is substantial" (Davis & Taylor, 1999, p. 69). Levesque and Gelles (1998) were the first to present a meta-analysis of 17 batterers' treatment outcome studies. Based on the small effect sizes ( $h$ s ranging from 0.18 to 0.27), they concluded that batterers' interventions "work a little, probably."

To help to clarify some of these discrepant conclusions, we conducted a formal meta-analysis, including the more methodologically rigorous studies, and new findings on recently completed experiments. The current study is the first formal meta-analysis on batterers' treatment outcome studies to be published to date. We attempted to improve on previous research in two ways. First, Hamberger and Hastings (1993) included studies that utilized uncontrolled, pre–post designs in their review. The level of confidence that any change in batterers' behavior was, indeed, due to treatment was undermined because extraneous causes were not ruled out by the presence of a control group. Pre–post studies preclude the estimate of an effect size due to treatment, as they are confounded with the effects of the legal system, i.e., the effects of "getting caught." As such, the present study utilized only studies that possessed some type of control group (e.g., treatment dropouts, another type of nonequivalent control group, or those randomly assigned to a no-treatment condition). Second, previous studies (Davis & Taylor, 1999; Levesque & Gelles, 1998;) have reported the effect size of batterers' treatment in terms of Cohen's  $h$  (Cohen, 1988). However, this statistic does not adjust for sample size and is more commonly used in power analysis than meta-analysis. To account for sample size, Cohen's  $d$  was selected as the measure of effect size in the present study.

## 1.2. Batterers' interventions

Only a few intervention modalities have been subjected to rigorous empirical test. These include feminist psychoeducational men's groups, cognitive–behavioral men's groups, anger management (a form of cognitive–behavioral group treatment), and couples' therapy.

### *1.2.1. Psychoeducational model*

The most prominent type of clinical intervention with batterers is a feminist psychoeducational approach (Pence & Paymar, 1993). This intervention, originated by the Duluth Domestic Abuse Intervention Project program in Minnesota, is frequently referred to as the Duluth model. According to this model, the primary cause of domestic violence is patriarchal ideology and the implicit or explicit societal sanctioning of men's use of power and control over women. This program, developed from a social work perspective, typically eschews DSM-type diagnoses and does not consider the intervention to be therapy. Rather, group facilitators lead consciousness-raising exercises to challenge the man's perceived right to control or dominate his partner. A fundamental tool of the Duluth model is the "Power and Control Wheel," which illustrates that violence is part of a pattern of behavior including intimidation, male privilege, isolation, emotional, and economic abuse, rather than isolated incidents of abuse or cyclical explosions of pent-up anger or painful feelings (Pence & Paymar, 1993). The treatment goals of the Duluth model are to help men change from using the behaviors on the Power and Control Wheel, which result in authoritarian and destructive relationships, to using the behaviors on the "Equality Wheel," which form the basis for egalitarian relationships (Pence & Paymar, 1993). The feminist Duluth-type model remains the unchallenged treatment of choice for most communities. In fact, the states of Iowa and Florida mandate that battering intervention programs adhere to the general tenets of the Duluth model to be state certified (Abel, *in press*; Healey, Smith, & O'Sullivan, 1998).

### *1.2.2. Cognitive behavioral groups*

An alternative to the feminist psychoeducational group is the cognitive-behavioral therapy (CBT) model. Cognitive behavioral batterers interventions, developed primarily by psychologists, tend to make violence the primary focus of treatment. Since violence is a learned behavior, nonviolence can similarly be learned according to the cognitive-behavioral model (Adams, 1988). Violence continues because it is functional for the user, reducing bodily tension, achieving victim compliance, putting a temporary end to an uncomfortable situation, and giving the abuser a sense of power and control (Sonkin, Martin, & Walker, 1985). Recognizing the functional aspects of violence, the cognitive-behavioral therapist points out the pros and cons of violence. In addition, they use skills training (e.g., communication, assertiveness, and social skills training) and anger management techniques (e.g., timeouts, relaxation training, and changing negative attributions) to promote awareness of alternatives to violence.

The intervention labels are often misleading. Some CBT groups are not strictly "cognitive" or "behavioral," as they address emotional components of violence, such as empathy and jealousy (Dunford, 2000). Most modern cognitive-behavior groups also usually address perpetrator attitudes and values regarding women and the use of violence toward women. To the extent that CBT groups address patriarchal attitudes, and Duluth model groups address the learned and reinforced aspects of violence, any distinction between CBT and Duluth model groups becomes increasingly unclear.

### 1.2.3. *Other modes of therapy*

The rationale for the use of group therapy is that men learn to confront one another's denial and victim blaming (Murphy & Baxter, 1997). As such, there have been no controlled, empirical studies to date testing individual therapy approaches for abusers. Due to concerns about the effectiveness of male-only group interventions, some in the domestic violence field are exploring alternatives to the psychoeducational group approach by testing conjoint groups (Dunford, 2000; O'Leary, Heyman, & Neidig, 1999). Advocates of couples groups state that including the wife in the group intervention may change the tenor of the men's group by rendering role-play more realistic and by reducing "women bashing" (Dunford, 2000). It may also empower the wife by allowing her to "witness authority figures confronting the offensive and oppressive nature of spouse abuse," as well as model for her constructive ways to deal with conflict (Dunford, 2000, p. 469). However, most states set standards, guidelines, or mandates that discourage or prohibit the funding of any program that offers couples or family counseling as a primary mode of intervention (Healy et al., 1998; Lipchick, Sirles, & Kubicki, 1997), as the woman's disclosures in the presence of her partner may lead to later retribution (Lipchick et al., 1997) or imply that she is at least partially to blame for her victimization (Jacobson, 1993).

## 2. Method

### 2.1. *Overview of methods of prior studies*

The primary purpose of this article is to quantitatively summarize the findings to date on the effect of batterers' treatment on violence recidivism. A review of the batterers' treatment literature was conducted using PsycInfo, entering the keywords "batterers" and "domestic violence." These were cross-referenced with terms including "treatment" and "intervention." Studies identified in this way were retrieved and their reference sections reviewed for additional treatment outcome studies. Additionally, the reference sections of five reviews of the batterer treatment literature were examined (Bargarozzi & Giddings, 1983; Davis & Taylor, 1999; Gelles, 1980; Rosenfeld, 1992; Tolman & Bennett, 1990). Prior to results of our quantitative meta-analysis, we will briefly summarize the methods and findings of available studies to the present, casting a broad net to include published materials, manuscripts in press, and data presented at national conferences. For three recent studies (Feder & Forde, 1999; Gondolf, 2000; Taft, Murphy, Elliott, & Morrel, 2001), additional information needed to calculate effect size was obtained directly from the authors.

#### 2.1.1. *Quasi-experimental studies*

Table 1 presents the quasi-experimental studies, most of which used the nonequivalent control group design (Campbell & Stanley, 1963) to compare either treatment completers to treatment dropouts or treated offenders to a matched group of nontreated batterers (not using true random assignment). It should be noted that the nonequivalent control group design employed by most studies on battering interventions does not meet the American Psycho-

Table 1  
Quasi-experimental designs

Study author	Group design and initial sample size	Treatment type	Treatment length	Attrition rates	Follow-up recidivism measure and response rates	% Re-offended	Effect size ( <i>d</i> )
Taft et al. (2001), Morrel, Elliott, Murphy, and Taft (2003), and Murphy (personal communication)	Tx1 completers ( <i>n</i> = 33); Tx2 completers ( <i>n</i> = 41), dropouts ( <i>n</i> = 12)	Tx1 = supportive + treatment retention; Tx2 = CBT + treatment retention	16 sessions	18% completed < 12 weeks	Police records at 22–36 months (73% of sample) and partner report (61% of the sample) at 6 months follow-up	Police report: Tx1 = 9.5%; Tx2 = 9.7%, dropouts = 54% Partner report: Tx1 = 10%; Tx2 = 18.5%, dropouts = 33%	Police report: Tx1 = 1.15; Tx2 = 1.22 Partner report: Tx1 = 0.69; Tx2 = 0.36
Gondolf (1997, 1998, 2000, personal communication)	Tx1 completers ( <i>n</i> = 158); Tx1 dropouts ( <i>n</i> = 55); Tx2 completers ( <i>n</i> = 145 ); Tx2 dropouts ( <i>n</i> = 64); Tx3 completers ( <i>n</i> = 140 ); Tx3 dropouts( <i>n</i> = 75); Tx4 completers ( <i>n</i> = 135); Tx4 dropouts( <i>n</i> = 72)	Four Duluth model programs of different lengths	Tx1 (Pittsburgh): 12 weeks with few additional services; Tx2 (Denver): 26 weeks; Tx3 (Houston): 24 weeks; Tx4 (Dallas): 12 weeks with several additional services	32% across all sites attended less than 2 months	Police reports (57%) at 15 months follow-up and cumulative partner, perpetrator, police report (48% of sample) at 30 months	Police report: Tx1 = 17%; Tx1 dropouts = 41%; Tx2 = 26%; Tx2 dropouts = 51%; Tx3 = NA; Tx3 dropouts = NA; Tx4 = 12%; Tx4 dropouts = 19% Partner report: Tx1 = 40%; Tx1 dropouts = 50%; Tx2 = 35%; Tx2 dropouts = 55%; Tx3 = 35%; Tx3 dropouts = 59%; Tx4 = 33%; Tx4 dropouts = 58%	Police report: Tx1 = 0.58 <sup>a</sup> ; Tx2 = 0.54; Tx3 = NA; Tx4 = 0.20 Partner report: Tx1: 0.20; Tx2: 0.41; Tx3: 0.50; Tx4: 0.52
Babcock and Steiner (1999)	Tx completers ( <i>n</i> = 106); Tx dropouts ( <i>n</i> = 178); incarcerated ( <i>n</i> = 55)	Multisite, majority Duluth model, psychoeducational + probation	36 weeks	68% completed < 28 sessions	Police report at 2 years postprosecution	Completers = 8%; dropouts = 23%; incarcerated = 62%	Tx vs. dropouts = .40

Murphy et al. (1998)	Tx completers ( $n = 10$ ); noncompleters ( $n = 225$ )	Duluth model psychoeducational	22 sessions	84% (of 62 men ordered to treatment) completed < full 22 weeks	Police records 12–18 months postprosecution	Completers = 0%; noncompleters = 16%	0.44
Dutton et al. (1997)	Tx completers ( $n = 156$ ); Tx dropouts and rejected ( $n = 290$ )	Clinical anger management vs. dropouts and rejected (for noncooperation, psychosis, etc.)	16 weeks	52%	Police reports ranging up to 11 years (mean 5.2 years)	Completers = 18%; dropouts = 21%	0.07
Dobash et al. (1996)	Tx completers ( $n = 40$ ); Tx dropouts ( $n = 80$ )	Psychoeducational group vs. dropouts	unknown	66%	Police and partner report (25% of sample) at 1 year follow-up	Police report: completers = 7%; dropouts = 10% Partner report: dropouts = 75%	Police report = 0.11 Partner report = 0.92
Newell (1994)	Tx1 = DV group completers ( $n = 155$ ); Tx1 dropouts ( $n = 118$ ); Tx2 = other Tx ( $n = 83$ ); no Tx ( $n = 135$ )	Feminist psychoeducational group vs. other Tx (AA, couples, individual) vs. group dropouts vs. no Tx control	12 weeks	57%	Police reports (re-arrest) at 2 year follow-up	Tx1 completers = 23%; Tx1 dropouts = 36%; Tx2 = 16%; no Tx = 22%	Tx1 completers vs. dropouts = 0.29; Tx1 completers vs. no Tx = -0.02 Tx2 vs. no Tx = 0.15
Flournoy (1993)	Tx1 ( $n = 16$ ); Tx2 ( $n = 13$ ); waitlist control ( $n = 14$ )	Tx1 = CBT; Tx2 = psychoeducational; control = waitlist	8 weeks	CBT 19%; psychoeducational 38%	Police reports 2–3 months follow-up (81% of sample)	Tx1 = 8%; Tx2 = 0%; control = 7%	Tx1 = -0.03; Tx2 = 0.33
Harrell (1991)	Tx1 ( $n = 81$ ); no-treatment control ( $n = 112$ )	Mandated CBT group (8–12 weeks) vs. no treatment mandated	8–12 weeks	20%	Police reports at 15–29 months; partner report on 90% of sample at 6 months	Police report: Tx = 50%; no Tx = 30% Partner report: Tx = 43%; no Tx = 12%	Police report = -0.42 Partner report = -0.76
Chen, Bersani, Myers, and Denton (1989)	Mandated to Tx ( $n = 120$ ); not mandated ( $n = 101$ )	Anger management	8 weeks	37% completed less than 7 sessions	Police reports	Completers = 5%; dropouts = 10%	0.19
Edleson and Grusznski (1988) Study 3	Tx completers ( $n = 84$ ); Tx dropouts ( $n = 37$ )	Psychoeducation followed by process oriented	8 weeks psychoeducation + 16 weeks	31%	Partner report at 6 months follow-up	Completers = 42%; dropouts = 49%	0.14

(continued on next page)

Table 1 (continued)

Study author	Group design and initial sample size	Treatment type	Treatment length	Attrition rates	Follow-up recidivism measure and response rates	% Re-offended	Effect size ( <i>d</i> )
Edleson and Grusznski (1988) Study 1	Tx completers ( <i>n</i> = 27); Tx dropouts ( <i>n</i> = 30)	Psychoeducation followed by process oriented	8 weeks psychoeducation + 16 weeks	47%	Partner report at 6 months follow-up	Completers = 33%; dropouts = 46%	0.26
Hamberger and Hastings (1988)	Tx completers ( <i>n</i> = 32); Tx dropouts ( <i>n</i> = 36)	CBT group	15 weeks	53%	Combination of self + partner + police report at 1 year follow-up	Completers = 9%; dropouts = 17%	0.23
Waldo (1988)	Tx completers ( <i>n</i> = 30); Tx dropouts ( <i>n</i> = 30); control ( <i>n</i> = 30)	Relationship enhancement men's group	12 weeks	50%	Police reports at 1 year follow-up	Completers = 0%; dropouts = 20%; controls = 20%	Completers vs. dropouts = 0.70; completers vs. control = 0.70
Leong, Coates, and Hoskins (1987)	Tx completers ( <i>n</i> ≈ 33); Tx dropouts ( <i>n</i> ≈ 34)	CBT group	unknown	≈ 50%	Police report at 6 months follow-up	Completers = 13%; dropouts = 29%	
Hawkins and Beauvais (1985)	Tx completers ( <i>n</i> = 52); Tx dropouts ( <i>n</i> = 43)	CBT	1–6 group + 6 couple and individual	45%	Police report at 6 months follow-up	Completers = 18%; dropouts = 18%	0.00
Stacey and Shupe (1984)	Initial <i>N</i> = 193; Tx1 ( <i>n</i> at follow-up = 77); dropouts ( <i>n</i> at follow-up = 30)	Multisite: 2 sites CBT, 1 site psychodynamic/Rogerian	10–18 weeks	Unknown	Partner report at 0–24 month follow-up (55% of sample)	Completers = 34%; dropouts = 50%	0.33

Tx = treatment.

<sup>a</sup> Effect sizes from the Pittsburgh site (Gondolf, 2000) were excluded from the meta-analysis due to treatment dropouts receiving additional legal sanctions.



logical Association's standards for establishing empirical support (Chambless et al., 1996). All of the quasi-experimental studies share the methodological problem of potentially "stacking the deck" in favor of treatment. Men who choose to complete treatment are known to be different from those who drop out (e.g., more educated, more likely to be employed, married, and Caucasian, and less likely to have a criminal record) (Babcock & Steiner, 1999; Hamberger & Hastings, 1988). Two studies did attempt to control for these preexisting group differences (Babcock & Steiner, 1999; Gondolf, 1997), and found that the effect attributable to treatment remained statistically significant. However, the percentages and effect sizes presented in Table 1 are not corrected for confounds due to group differences between treatment dropouts and completers. It is difficult to estimate the effect size controlling for demographic variables because most studies do not present the data in a manner such that a reanalysis, controlling for confounds, would be possible; we predict that doing so would dramatically decrease the effect size.

A second methodological difficulty with quasi-experimental designs is the degree to which the "dropout" condition is contaminated by the effects of alternative sanctions against batterers. Gondolf (2000) found that the effect size of treatment at one of his study's site (Pittsburgh) was negligible. He later learned that alternative sanctions were issued upon treatment dropouts, rendering the dropouts invalid as a comparison group. Thus, that site was excluded from our meta-analysis. Other studies that were included in the meta-analysis may have similar confounds that are undisclosed in their reports. In some cases, quasi-experimental designs and randomized experiments can yield comparable effect sizes (Heinsman & Shadish, 1996). However, whether this is true for the body of studies on batterers' intervention remains an empirical question.

### 2.1.2. *True experiments*

Because of the ever-present risk of confounds among quasi-experimental studies, results from randomized experiments are the "gold standard" for meta-analyses (Shadish & Ragsdale, 1996, p. 1290). Therefore, results of the five, recent experimental studies should be considered a more accurate estimate of the actual effect size due to batterers' treatment. Table 2 presents the five studies to date that have employed random assignment. These five experiments deserve special attention.

Feder and Forde (1999) randomly assigned batterers on probation to either a feminist-psychoeducational program or no treatment in Broward County, FL. In general, there were no statistically significant differences between the two groups on recidivism as measured by police records ( $d=0.04$ ) or by victim report ( $d=-0.02$ ). There was a small but significant effect on recidivism among the subset of men randomly assigned to group treatment who attended all 26 sessions. In this study, random assignment apparently failed, with an uneven number of men being assigned to the treatment and control condition (Feder & Forde, 1999). Moreover, this study suffered from a particularly high attrition rate of men from treatment (60%) and low response rate from victims at follow-up (22%).

In a large evaluation of U.S. Navy personnel stationed in San Diego, Dunford (1998, 2000) compared a 36-week cognitive-behavioral group and a 26-week couples therapy format to a rigorous monitoring condition and a no-treatment control (victims safety planning). Neither

Table 2  
Experimental designs

Study authors	Group design and initial sample size	Treatment type	Treatment length	Attrition rates	Follow-up recidivism measure and response rates	% Re-offended	Effect size ( <i>d</i> )
Feder and Forde (1999, personal communication)	Tx = Duluth ( <i>n</i> = 174); control ( <i>n</i> = 230)	Duluth + probation vs. probation only	26 weeks	60%	Police at 1 year and partner report (22% of sample) at 6 month follow-up	Police report: Tx = 4.8%; control = 5.7% Partner report: Tx = 32.7%; control = 31.6%	Police report: Tx = 0.04 Partner report: Tx = -0.02
Dunford (2000)	Tx1 = CBT ( <i>n</i> = 168); Tx2 = couples ( <i>n</i> = 153); monitoring ( <i>n</i> = 173); control ( <i>n</i> = 150)	CBT men's group, conjoint Tx, and rigorous monitoring vs. victim safety planning control	Tx1 = 36 weeks + 6 monthly meetings; Tx2 = 26 weeks + 6 monthly meetings; monitoring = monthly meetings for 12 months	29%	Police and partner report (72% of initial sample of 861) on at 1 year follow-up	Police report: Tx1 = 4%; Tx2 = 3%; monitoring = 6%; control = 4% Partner report: Tx1 = 29%; Tx2 = 30%; monitoring = 27%; control = 35%	Police report: Tx1 = 0.00; Tx2 = 0.05; monitoring = -0.09 <sup>a</sup> Partner report: Tx1 = 0.13; Tx2 = 0.10; monitoring = 0.17 <sup>a</sup>

Davis et al. (2001)	Tx1 = long ( $n = 129$ ); Tx2 = brief ( $n = 61$ ); control ( $n = 186$ )	Duluth model treatments vs. community service control	Tx1 = 26 weeks; Tx2 = 8 weeks	33%	Police and partner report of new incident in past 2 months (50% of sample) at 1 year follow-up	Police report: Tx1 = 10%; Tx2 = 25%; control = 26% Partner report: Tx1 = 14%; Tx2 = 18%; control = 22%	Police report: Tx1 = 0.41; Tx2 = 0.02 Partner report: Tx1 = 0.21; Tx2 = 0.10
Ford and Regoli (1993)	Tx1 = pretrial diversion into counseling ( $n = 127$ ); Tx2 = counseling as condition of probation ( $n = 114$ ); control = sentence without counseling ( $n = 106$ )	Counseling (unknown type) as pretrial diversion vs. condition of probation vs. other sentencing (e.g., fine, jail) control	Unknown	Unknown	Partner report at 6 month follow-up (31% of sample)	Tx1 = 34%; Tx2 = 45%; control = 34%	Tx1 = 0.00; Tx2 = - 0.22
Palmer et al. (1992)	Tx ( $n = 30$ ); control ( $n = 26$ )	Psychoeducational vs. probation only	10 weeks	30% attended < 7 sessions	Police at 1–2 year follow-up	Tx = 10%; control = 31%	Tx = 0.54

Tx = treatment.

<sup>a</sup> Effect sizes generated from the rigorous monitoring conditioning (Dunford, 2000) were excluded from this meta-analysis, as it does not represent a therapeutic intervention. Weighted percentage of nontreated who re-offended based on police report = 21%; based on partner report = 35%.

CBT men's groups ( $d=0.13$ ) nor couples therapy ( $d=0.10$ ) had a significant impact on recidivism at 1-year follow-up based on victims' report. This study represents the most methodologically rigorous study conducted to date in terms of sample size, length of follow-up, attrition rates, follow-up reporting rates, and assessment of treatment adherence. However, it is important to note that this sample of batterers, those employed through the Navy in San Diego, are not representative of the population of batterers court-mandated to domestic violence programs around the country. All of the research participants were employed, had a high stake in social conformity, and thus, were more "socially bonded" (Sherman, Smith, Schmidt, & Rogan, 1992). Any intervention, including arrest and being identified by authorities, may work to deter socially bonded individuals from repeat offenses. This may be reflected in the unusually low official recidivism rates of the nontreated batterers (4%).

Davis, Taylor, and Maxwell (2001) compared a long (26-week) psychoeducational group to a brief (8-week), psychoeducational group, and to a community service control (70 hours of clearing vacant lots, painting senior citizen centers, etc.) in Brooklyn, NY. They found a statistically significant reduction in recidivism and a small but respectable effect size of  $d=0.41$  based on criminal records among the long treatment group only; the 8-week group was indistinguishable from the community service control ( $d=0.02$ ). As for partner report, this study employed a rather unusual method of calculating re-offenses. Only new incidents of violence in the 2 months prior to the follow-up contact point were included rather than a cumulative count. When based on victim report of these recent offenses, neither the long nor the brief intervention had a statistically significant effect on re-assault when compared to no treatment. Correspondingly, the effect size due to treatment based on partner report of subsequent violence was small ( $d=0.21$ ). It is important to note that, like in the Broward County experiment (Feder & Forde, 1999), random assignment may have been compromised. In the Brooklyn experiment (Davis et al., 1998), nearly 30% of initial assignments were subjected to "judicial overrides" (Gondolf, 2001); that is, judges reassigned defendants to different interventions.

Ford and Regoli (1993) designed a study that randomly assigned batterers into treatment as a pretrial diversion (i.e., defendants' criminal records would be cleared pending treatment completion), treatment as a condition of probation postconviction, vs. alternative sentencing strategies (e.g., paying a fine or going to jail). Although this study was designed to test different sentencing options rather the effects due to treatment, one can compare batterers sentenced to treatment vs. batterers not sentenced to treatment (although the type of treatment and actual attendance rates were not specified). Again, there were no significant differences or effect sizes comparing recidivism rates based on victim report between men sentenced to treatment vs. those who were not. Neither treatment as pretrial diversion ( $d=0.00$ ) nor as a condition of probation postconviction ( $d=-0.22$ ) was found to be superior to purely legal interventions.

Finally, Palmer, Brown, and Barrera (1992) conducted a small scale study in Canada of men using block random procedure: men were assigned to 10-week psychoeducational treatment if a new group was to commence within 3 weeks or, if not, to a "probation only" control group. The relatively unstructured, client-centered treatment addressed beliefs about

violence, responsibility for violent behavior, coping with conflict and anger, self-esteem, and relationships with women (Peluch, 1987). Based on police reports, men assigned to the treatment condition re-offended at a significantly lower rate than men assigned to probation only, yielding a medium effect size ( $d=0.54$ ). However, this study is limited by its small sample size, and the results may not be generalizable to other samples.

Conducting an experiment in which judicial discretion is sacrificed and criminals are randomly assigned to treatment or no treatment can be problematic on ethical as well as practical grounds (Dutton, Bodnarchuk, Kropp, & Hart, 1997). Adopting an experimental design does not guarantee a more rigorous evaluation than quasi-experimental designs afford (Gondolf, 2001). While it is true that experimental designs permit greater confidence in conclusions regarding causal relations, it is also the case that problems with differential attrition and failure of random assignment reduce internal validity of this design. Additionally, researchers must grapple with the “intention-to-treat” problem: should effect sizes be calculated from the initial sample size or from the completers only? What if the majority of “treated” offenders attended no-treatment groups whatsoever? It is recommended that researchers report both recidivism rates for all batterers who were assigned to treatment as well as those who actually completed treatment (although few of studies have done so).

## 2.2. Study inclusion criteria

Originally, 68 empirical studies of the efficacy of batterers’ treatment programs were located. These studies were classified according to the design: experimental ( $k=5$ ), quasi-experimental ( $k=17$ ) and pre–post ( $k=48$ ). The criterion for inclusion in this meta-analysis was the (1) the presence of some form of comparison group of batterers and (2) reliance on victim report or police record as the index of recidivism. The uncontrolled, pre–post test studies have been reviewed previously (Davis & Taylor, 1999; Hamberger & Hastings, 1993; Rosenfeld, 1992). These are the weakest methodological designs and generally tend to overestimate effect size (Lipsey & Wilson, 1993). On this basis, 48 of the 70 studies were not included. The stronger quantitative evaluations of domestic violence interventions generally fall into two categories: (1) quasi-experimental, where treatment completers are compared to treatment dropouts or to a matched comparison group that did not receive treatment and (2) true experimental designs, where clients are randomly assigned to treatment(s) vs. no treatment. Studies ( $k=22$ ) consisting of experimental and quasi-experimental designs formed the data for this quantitative review. These studies yielded a total of 44 effect sizes (effect sizes formed the unit of analysis for the present study), in which a treatment group was compared to either a randomized control or treatment dropouts.

Several recent studies have compared two active treatments for domestic violence without the inclusion of a control group and without the comparing of completers to a no-treatment or dropout comparison group. For example, two studies (Brannen & Rubin, 1996; O’Leary et al., 1999) compared a couples format to a gender-specific group format. Saunders (1996) compared the response of batterers with different personality profiles to both more standard structured groups vs. more process-oriented therapies. Although one could calculate an effect size from these treatment comparison studies, the resultant statistic would reflect the

magnitude of the difference between two active treatments. This would grossly underestimate the effect size due to treatment and potentially bias the results against treatment. Other studies (e.g., Edleson & Syers, 1991) report recidivism rates for the treatment completers only. Sullivan and Bybee (1999) conducted a well-designed randomized clinical trial of victims' advocacy intervention that yielded an effect size ( $d$ ) of 0.35 based on victim report of recidivism; however, interventions with victims are beyond the scope of the current meta-analysis. While these studies are important contributions to the clinical literature, they do not render effect sizes estimating the effect due to batterers' treatment. As such, they are not included in this meta-analysis.

### 2.3. Coding

Effect sizes were coded along a number of dimensions that were theoretically promising for investigation as moderators. Each effect size was classified according to the type of report upon which recidivism rates were based, treatment type, treatment duration, follow-up time, and attrition rates. Report of batterer recidivism took two forms: police report and partner report. Many of the earlier effect sizes relied exclusively on batterers' self-report as an outcome measure. Such effects cannot differentiate between treatment success and batterers' tendency to vastly underreport the true incidence of abuse (Davis & Taylor, 1999; Rosenfeld, 1992). Moffit et al. (1997) found that the reliabilities between male and female report in a community sample on the presence/absence of violence was poor (average kappa=.36). They concluded that in a therapeutic or correctional setting, "where the pressures (for batterers) to bias their reports may outweigh researchers' promises of confidentiality," collateral reports may be essential (Moffit et al., 1997, p. 54). In light of this potential reporting bias, only effect sizes that use at least one independent report of recidivism, either victim report or criminal record, were included in this review. In many cases, both police and partner reports were examined. As such, our analyses of the data examined separate effect sizes for all comparisons presented in each study; moderator analysis was performed twice, separately for partner and police report, to avoid artificial inflation of the number of studies.

#### 2.3.1. Recidivism

Considering practical significance, most clinicians working with batterers agree that cessation of intimate partner violence is an important success criterion (Edleson, 1996), rather than, for example, showing a decrease in the frequency or severity of violent acts. For the purposes of this review, "recidivism" is considered any report of physical violence reported by the victims and/or any domestic violence incidents reported to the police during a follow-up period (i.e., recidivism is a dichotomous variable and the effect sizes are calculated from the proportion of men who re-offended). Most studies utilized the Conflict Tactics Scale (CTS/CTS-2) (Straus, 1979; Straus, Hamby, Boney-McCoy, & Sugarman, 1996) as an outcome measure of partner report of recidivism. Our dependent measure of partner report of recidivism was the percentage that stopped further physical aggression, rather than change in CTS score. While a cessation of men's emotional abuse and increased accountability are

also goals of many treatment programs, few program evaluations or experiments have examined any outcome variable other than physical abuse frequency or criminal recidivism. Therefore, only police report and partner reports of physical assault were included in this meta-analytic review.

### 2.3.2. *Treatment types*

Treatment types included Duluth/feminist psychoeducational ( $k=19$ ), CBT ( $k=11$ ), and other ( $k=7$ ). Since the meta-analytic method requires examination of the heterogeneity/homogeneity of variance due to each putative moderator, a cell size greater than one was required for each level of the moderator. As a consequence, any type of treatment that occurred only once in the database was aggregated into the “other” category. The seven effect sizes in the “other” category came from studies testing the effectiveness of couples therapy (Dunford, 2000), supportive therapy (Taft et al., 2001), relationship enhancement (Waldo, 1988), a mixture of different interventions (Newell, 1994), and therapies of an unspecified type (Ford & Regoli, 1993).

### 2.3.3. *Treatment length*

Treatment length was dichotomized: short (mean treatment length <16 weeks), and long (mean treatment length 16 weeks). If any treatment did not maintain a uniform duration, the average length of treatment was utilized.

### 2.3.4. *Follow-up length*

Effect sizes were classified into one of two categories based on follow-up length: short (mean follow-up time <12 months) and long (mean follow-up time >12 months). For studies with variable follow-up times, the mean follow-up time was calculated.

### 2.3.5. *Attrition*

Attrition from treatment was calculated as the percentage of individuals who were classified as “dropouts” from the quasi-experimental studies by the authors. It should be noted that different authors have distinct criteria for what constitutes treatment completion. For some effect sizes, completers must attend 100% of the sessions; other authors report “completers” as those attending 80% or more of the required sessions. Due to the inconsistencies in calculating and reporting attrition, this variable was not entered into the meta-analysis. However, attrition rates are reported in [Tables 1 and 2](#). They may be viewed as an index of quality of treatment or quality of the coordinated community response and may influence the effect size.

## 2.4. *Reliability*

For reliability purposes, both the first and second authors reviewed and coded each study. There were no disagreements on study design, type of report, length of treatment, or follow-up length (reliability = 100% agreement). However, there was one study in which the coders disagreed on the treatment type (reliability = 95% agreement). In this case, the study author

was contacted (Jeffrey Edelson, personal communication, September 13, 2000) to assist in assigning a label to the treatment.

### 2.5. Estimates of effect size

Table 1 presents the general design, type of treatment, and recidivism or re-offense rates of all identifiable quasi-experimental designs, and Table 2 the existing true experimental studies conducted in the past decade. The re-offense rates (that is, the percentage in the treated and control conditions who re-offended) as reported in the studies were then recalculated into an effect size, using the  $g$  statistic on proportions (Hedges & Olkin, 1985). The  $g$  statistic on proportions was then transformed into the  $d$  statistic, adjusting for sample size (Johnson, 1995). It is important to note that the size of the final samples with complete recidivism data at follow-up, especially those based on partner/victim report, is usually significantly smaller than the initial  $n$ . In many cases, the specific  $n$ s of treated and comparison groups with complete follow-up data were not explicit, although the follow-up response rate usually was. In those cases, we estimated the final  $n$  by “discounting” the initial  $n$  in each condition by the proportion with complete follow-up data.

An “effect size” is an attempt to quantify the magnitude of the effect due to treatment using a shared metric than is not influenced by the size of the sample. When based on the  $d$  statistic, effect sizes of 0.20 are considered “small,” 0.50 are considered “medium,” and effect sizes 0.80 and above are considered large (Cohen, 1988). The  $d$  effect size is in units of standard deviations; therefore, an effect size of 0.25 translates to an improvement of one-fourth of a standard deviation compared to no treatment. In true experimental designs, the effect size allows us to evaluate the magnitude of the impact that treatment has on recidivism; in quasi-experimental designs, the effect size approximates the strength of relationship between treatment and recidivism, uncontrolled for external confounds (Campbell & Stanley, 1963).

Effect sizes and variances were calculated in terms of  $d$  using Hedges and Olkin’s (1985) meta-analytic method. This enabled differential weighting of effects for sample size. Calculation of the  $d$  was accomplished utilizing D-Stat version 1.11 (Johnson, 1995). This software program calculates  $d$  based on proportions by treating each proportion as the mean of a distribution of successes vs. failures. Effect sizes were computed for each comparison for each dependent measure (i.e., report type), resulting in a total of 37 effect sizes. Moderator analysis was then conducted using MetaWin 1.0 (Rosenberg, Adams, & Gurevitch, 1997). This computer program follows Hedges and Olkin’s hierarchical approach to meta-analysis that employs the  $Q$  statistic to determine the degree of heterogeneity that exists between and within groups. As mentioned previously, other studies (Davis & Taylor, 1999; Levesque & Gelles, 1998) have reported the effect size of batterers’ treatment in terms of Cohen’s  $h$  (Cohen, 1988). Recalculating the effect sizes in terms of Cohen’s  $h$  does not substantially change the conclusions of this article. The  $d$  effect sizes can easily be converted to  $r$  effect sizes (Wolf, 1986, p. 35)<sup>1</sup> to calculate a

<sup>1</sup> Formula for  $r$ -to- $d$  transformation:  $r = \frac{d}{\sqrt{d^2 + 4}}$ .



binomial effect size display (BESD), using the formula (Rosenthal, 1995; Rosnow & Rosenthal, 1988):

$$\text{BESD} = 0.50 + (r/2)$$

The BESD allows for translation of the effect size in terms of differential outcome rates to assist in interpreting the practical importance of the effect size.

Previous works (Babcock & LaTaillade, 2000; Davis & Taylor, 1999) have informally examined the effect of batterers' treatment by taking the average effect size across study. In contrast, formal meta-analyses weight effect sizes by sample size. Therefore, the results of this article may differ substantially from simply averaging or “eyeballing” of the effect sizes presented in the tables.

### 3. Results

Based on the data summarized in Table 1, the weighted percentage of nontreated offenders who recidivated was 21% based on police reports and 35% based on partner reports. These recidivism rates for nontreated offenders are consistent with those previously reported (O'Leary et al., 1989; Rosenfeld, 1992).

#### 3.1. Publication bias

Analysis for publication bias and the “file drawer” phenomenon was conducted using a normal-quantile plot (Wang & Bushman, 1998). If null findings were selectively ignored, the normal-quantile plot would reveal absence of effect sizes around zero. Examination of the plots revealed no evidence for a publication bias (see Fig. 1).

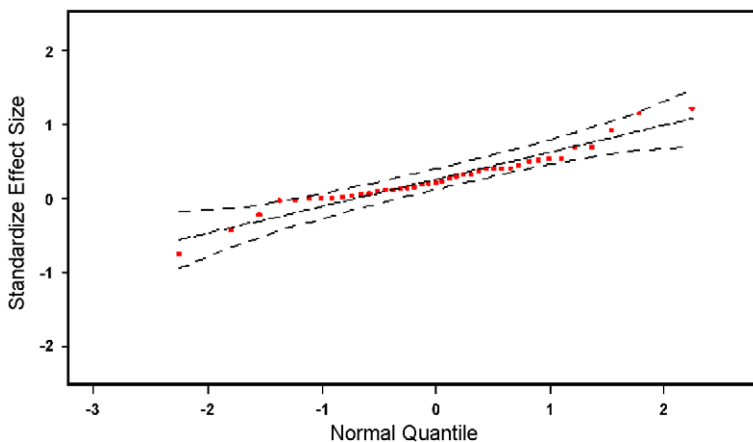


Fig. 1. Normal quantile plot to assess for the “file-drawer” problem.

### 3.2. Outlier analysis

Outlier analysis was conducted using the sample adjusted meta-analytic deviancy statistic (SAMD) developed by Huffcutt and Arthur (1995). The SAMD statistics were calculated separately for police and partner report. Examination of the scree plot of SAMD statistics when recidivism was assessed by police report suggested four possible outliers: both CBT (SAMD=8.73) and supportive interventions (SAMD=6.99) with retention techniques reported by Taft et al. (2001) and CBT in Harrell (1991) (SAMD=-11.08). Taft et al. and Harrell were thus excluded from subsequent analyses.

The scree plot of SAMD statistics based on partner report indicated that there were two outliers. These data points represented Dobash, Dobash, Cavanagh, and Lewis (1996) and Harrell (1991) with SAMDs of 11.01 and -15.02, respectively. Both effect sizes were excluded from the subsequent analysis based on outlier analysis.

### 3.3. Moderators of effect size

The remaining 36 effect sizes were entered into the hierarchical fixed effects analysis described by Hedges and Olkin (1985). A model was tested that reflected a combination of methodological and treatment moderators (Fig. 2); these included: report type, experimental vs. quasi-experimental design, and treatment type.

### 3.4. Effects due to method of assessing recidivism

The first moderator variable entered into the analysis was report type. The resulting analysis of two sets of effect sizes based on police and partner reports (i.e., a hierarchical moderator approach) permitted optimal use of the existing data without redundant use of samples in each group. We report 95% confidence intervals (CIs) for all effect size estimates. CIs that do not contain zero can be considered statistically significant from zero at the  $P < .05$  level. Effects based on police report ( $k=20$ ) yielded an overall effect size of  $d=0.18$  (95% CI=0.11–0.25) and the effects based on partner report ( $k=16$ ) yielded an equivalent effect size of  $d=0.18$  (95% CI=0.08–0.28). Examination of the  $Q$ -within statistic was not significant heterogeneity for police report ( $Q_w=26.96$ ,  $df=19$ , ns) or partner report ( $Q_w=10.96$ ,  $df=16$ , ns). A significant  $Q_w$  statistic indicates heterogeneity among the effect sizes that suggested the existence of further moderators. While the  $Q_w$  was not statistically significant for either police or partner report, indicating a lack of heterogeneity, the presence of the hypothesized model (Fig. 2) warranted continued examination of the remaining moderators (Rosenthal, 1995).

### 3.5. Effect due to study design

The second moderator variable entered into the model was research design (i.e., experimental or quasi-experimental). This variable was examined for effects based on police and partner report. Analysis of research design as a moderator for effect size within police report revealed that experimental designs ( $k=6$ ) had an overall  $d=0.12$  (95% CI=0.02–0.22). The

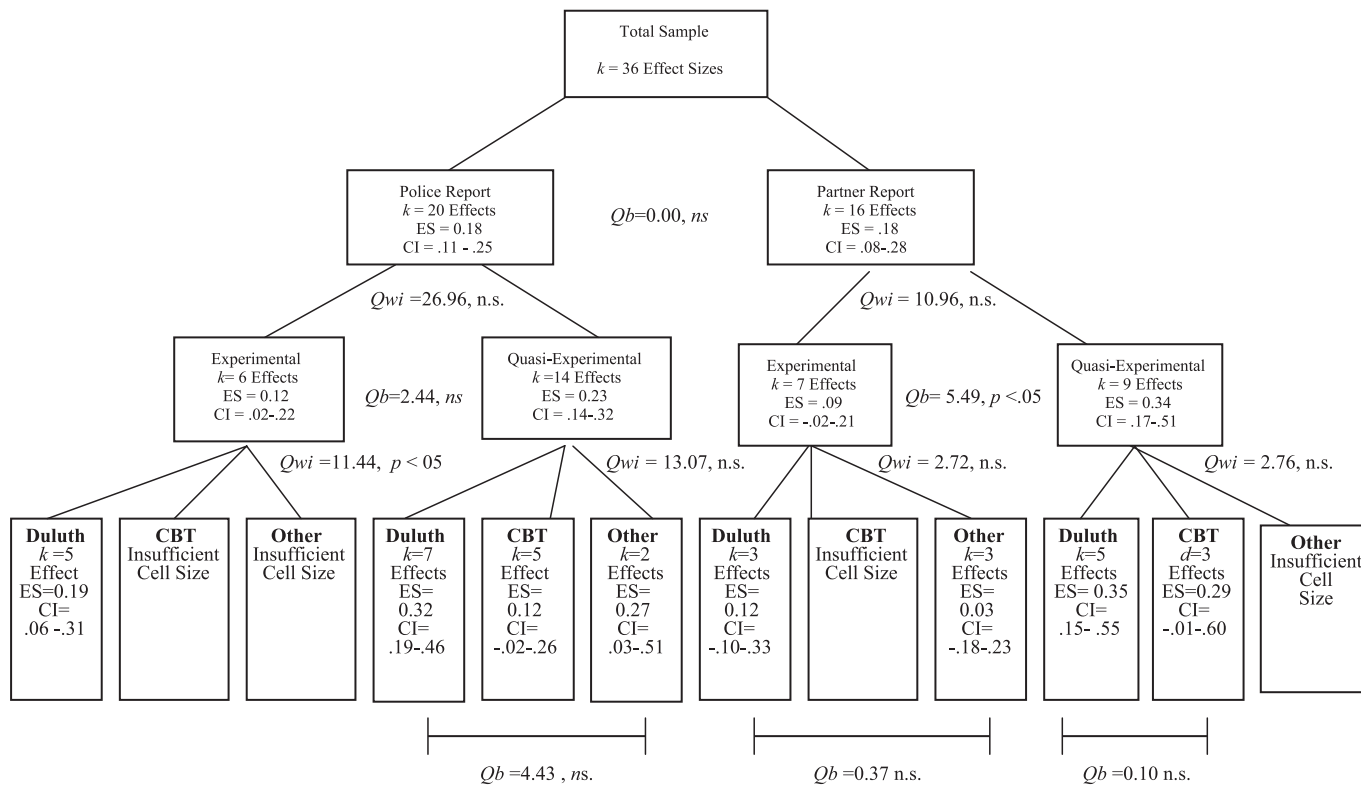


Fig. 2. Meta-analytic model testing recidivism report, study design, and type of treatment as moderators.

overall effect size for quasi-experimental designs with police report ( $k = 14$ ) was  $d = 0.23$  (95% CI = 0.14–0.32). For both experimental and quasi-experimental designs, treatment had a significant yet small impact on the cessation of domestic assault. There was not a significant difference between the overall effect sizes for experimental and quasi-experimental designs based on police report ( $Q_b = 2.44$ ,  $df = 1$ , ns). Examination of results based on police report indicated that there was significant heterogeneity among effect sizes among experimental designs ( $Q_w = 11.44$ ,  $df = 5$ ,  $P < .05$ .) but not for quasi-experimental designs ( $Q_w = 13.07$ ,  $df = 13$ , ns).

Similar analyses were conducted for effect sizes based on partner report ( $k = 16$ ). Analysis of research design as a moderator for effect size within partner report revealed an average effect size for experimental designs ( $k = 7$ ) of 0.09 (95% CI =  $-0.02$ – $0.21$ ), not significantly different from zero. The overall effect size based quasi-experimental designs with partner report ( $k = 9$ ) was  $d = 0.34$  (95% CI = 0.17–0.51). This represents a significant yet small effect size. There was a statistically significant difference between the overall effect sizes for experimental and quasi-experimental designs based on partner report ( $Q_b = 5.49$ ,  $df = 1$ ,  $P < .05$ .) Examination of the effect sizes based on experimental designs and partner report indicated that there was not significant heterogeneity ( $Q_w = 2.72$ ,  $df = 6$ , ns). Inspection of the effect sizes based on quasi-experimental designs and partner report indicates that there is not significant heterogeneity ( $Q_w = 2.76$ ,  $df = 8$ , ns) within these cells.

### 3.6. Effect due to treatment type

The third moderator variable entered into the model was treatment type. This was examined for effect sizes based on experimental design and police report, quasi-experimental design and police report, and quasi-experimental design and partner report. Calculation of the effect overall effect size due to treatment type within experimental designs with police report indicated that Duluth ( $k = 5$ ) had an effect size of  $d = 0.19$  (95% CI = 0.06–0.31). CBT and “other” therapies lacked sufficient cell size ( $k < 2$ ) to calculate an effect size. Thus, Duluth demonstrated a small effect based on police report and experimental design.

Examination of the effect overall effect size due to treatment type within quasi-experimental designs with police report indicated that Duluth ( $k = 7$ ) had an effect size of  $d = 0.32$  (95% CI = 0.19–0.46), CBT ( $k = 5$ ) had an effect size of  $d = 0.12$  (95% CI =  $-0.02$ – $0.26$ ), and other ( $k = 2$ ) had an effect size of  $d = 0.27$  (95% CI = 0.03–0.51). In this case, the effect sizes from the Duluth model and “other” interventions were significantly different than zero, whereas CBT interventions were not significantly different from zero. However, these effect sizes did not differ significantly from one another ( $Q_b = 4.43$ ,  $df = 2$ , ns).

Examination of the overall effect size due to treatment type within experimental designs with partner report indicated that Duluth ( $k = 3$ ) had an effect size of  $d = 0.12$  (95% CI =  $-0.10$ – $0.33$ ) and other ( $k = 3$ ) had an effect size of  $d = 0.03$  (95% CI =  $-0.18$ – $0.23$ ). CBT therapies lacked sufficient cell size ( $k < 2$ ) to calculate an effect size. Effect sizes did not differ significantly from one another ( $Q_b = 0.37$ ,  $df = 2$ , ns).

Calculation of the overall effect size due to treatment type within quasi-experimental designs with partner report indicated that Duluth interventions ( $k = 5$ ) had an effect size of  $d = 0.35$  (95%

Table 3  
Effect size due to factors not tested in the model

	<i>d</i> based on police report	<i>k</i>	95% CI	<i>d</i> based on partner report	<i>k</i>	95% CI
Overall treatment length	0.18	18	0.11–0.25	0.20	14	0.10–0.32
Long	0.16	8	0.08–0.25	0.18	8	0.06–0.31
Short	0.20	10	0.09–0.32	0.30	6	0.08–0.51
Overall follow-up length	0.18	19	0.11–0.25	0.18	16	0.08–0.28
Long	0.25	8	0.14–0.35	0.48	3	0.21–0.75
Short	0.13	11	0.04–0.22	0.13	13	0.02–0.24
Overall attrition rate	0.18	19	0.11–0.25	0.20	14	0.09–0.31
High ( $\geq 50\%$ )	0.20	8	0.10–0.30	0.09	2	–0.23–0.40
Low ( $< 50\%$ )	0.16	11	0.07–0.26	0.22	11	0.10–0.34

*k* = number of studies.

CI = 0.15–0.55) and CBT ( $k = 3$ ) had an effect size of  $d = 0.29$  (95% CI = –0.01–0.60), while the “other” category lacked sufficient cell size to be included in this analysis. The two effect sizes, however, were not significantly different from each other ( $Q_b = 0.10$ ,  $df = 1$ , ns).

Due to the small cell sizes for each treatment type a second model was examined that aggregated the experimental and quasi-experimental effect sizes for each reporting method (i.e., police or partner).  $Q$ -within and  $Q$ -between statistics were identical to the initial model for police and partner report. Calculation of the overall effect size due to treatment type within police report indicated that Duluth ( $k = 11$ ) had an effect size of  $d = 0.25$  (95% CI = 0.16–0.34), CBT ( $k = 6$ ) had an effect size of  $d = 0.09$  (95% CI = –0.03–0.20), and other ( $k = 3$ ) had an effect size of  $d = 0.09$  (95% CI = –0.01–0.32). There were no statistically significant differences between effect sizes for among the three treatment categories ( $Q_b = 4.80$ ,  $df = 2$ , ns).

Examination of the overall effect size due to treatment type within partner report indicated that Duluth ( $k = 8$ ) had an effect size of  $d = 0.24$  (95% CI = 0.09–0.39), CBT ( $k = 4$ ) had an effect size of  $d = 0.20$  (95% CI = –0.001–0.40), and other ( $k = 4$ ) had an effect size of  $d = 0.04$  (95% CI = –0.16–0.25). There were no statistically significant differences between effect sizes for among the three treatment categories ( $Q_b = 2.36$ ,  $df = 2$ , ns).

There was inadequate power to assess effect due to treatment length or follow-up length as moderator variables under different types of treatment. The overall effect sizes for treatment length, follow length, and attrition are reported in Table 3. Further analysis was conducted to examine the degree to which the inclusion of outliers in the analysis altered the present findings. In particular, a “best-case” scenario was evaluated in which only the low outliers were excluded from the analysis. The results were not significantly different from the model with all outliers removed.

#### 4. Discussion

In general, the effect size due to group battering intervention on recidivism of domestic violence is in the “small” range. There were no significant differences in average effect size

between Duluth-type and cognitive-behavioral battering intervention programs using either police records or victim reports as the index of recidivism. While quasi-experimental designs tended to yield higher effect sizes than true experiments, the differences in effect sizes were not significant. Regardless of reporting method, study design, and type of treatment, the effect on recidivism rates remains in the small range. In the best case scenario, using quasi-experimental designs based on partner report, the effect size is  $d=0.34$  indicating that treated offenders showed a one-third standard deviation in improvement in recidivism as compared to nontreated batterers.

If one relies exclusively on the five experimental studies, the effect sizes are even smaller. However, the effect sizes may be small as a result of measurement error and methodological difficulties common to research in applied settings (McCartney & Rosenthal, 2000). McCartney and Rosenthal (2000, p. 178) warn that “(g)iven that the stakes are so high, we should be wary of accepting the null hypothesis when it might very well be false—as it almost always is.” Based on the experimental studies, the effect size ( $d$ ) due to treatment is 0.09 and 0.12, based on victim report and police records, respectively. This means that treatment is responsible for an approximately one-tenth of a standard deviation improvement in recidivism. Based on a partner report, treated batterers have a 40% chance of being successfully nonviolent, and without treatment, men have a 35% chance of maintaining nonviolence. Thus, there is a 5% increase in success rate attributable to treatment. 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. Whether this success rate is cause for celebration or despair depends on a cost-benefit analysis; taking into account the cost of treatment and any potential “side effects” vs. the benefits of injury prevention and decreased psychological risk to the victim as well as the children exposed to family violence. A 5% decrease in violence may appear insignificant; however, batterers treatment in all reported cases of domestic violence in the United States would equate to approximately 42,000 women per year no longer being battered.

#### *4.1. How large of an effect size should we expect?*

One way to contextualize the effect size due to treatment is by comparing it to the effect sizes for treatment in other populations. Davis and Taylor (1999) compared their treatment effect size of 0.41 to the effect size of an early clinical trial on the effect of aspirin on heart attacks, which was only 0.068 and constitutes a 4% reduction in heart attacks (Rosnow & Rosenthal, 1988). Compared to this standard, they conclude that “the effect sizes seen in batterers’ treatment studies are quite substantial” (Davis & Taylor, 1999, p. 85). However, the average effect size across psychotherapy studies is much larger, approximately  $d=0.85$  (Smith, Glass, & Miller, 1980). In practical terms, psychotherapy leads to benefits in 70% of cases (Rosenthal, 1995). Compared to this standard, there is great room for improvement in our batterers’ treatment interventions.

However, comparison with psychotherapy outcomes in general may not be fair. Most psychotherapies address internalizing problems (e.g., depression, anxiety) rather than

externalizing problem behavior, like aggression. Given that aggression is difficult to treat, compounded with the fact that batterers are generally not seeking treatment voluntarily and do not necessarily expect the interventions to help (Gondolf, 2001), perhaps an overall small effect due to treatment is to be anticipated. A recent meta-analysis of psychotherapy with children and adolescents reveals that the effect size for treatments of aggression was  $d=0.32$  (Weisz, Weiss, Han, Granger, & Morton, 1995), indicating a 16% improvement in success rate over no treatment. Correctional treatments with adult prisoners result in effect sizes averaging  $d=0.25$  (Loesel & Koefler, 1987, cited in Lipsey & Wilson, 1993), approximating a 12% improvement rate. Based on Rosenfeld's (1992) earlier review of the literature, Dutton (1998, p. 177) speculated that the effects of battering interventions fall midrange between the effects due to psychotherapy and the effects due to rehabilitation of offenders. Results from this meta-analysis reveal that even Dutton's rather modest claim appears to be overly optimistic. The effects due to battering intervention are much closer to rehabilitation effects than the effect sizes of psychotherapy in general.

#### 4.2. Have "all won and all must have prizes?"

While the effect attributable to treatment is to some extent dependent on the methodologies employed by the studies, the effect sizes for Duluth model and CBT treatments remain relatively similar. With liberal estimates based on quasi-experimental studies, Duluth interventions yield a small effect size of  $d=0.35$  while CBT interventions yield a smaller effect size of  $d=0.29$ . Given the variability in effect sizes of the studies that make up these averages, however, we cannot say that CBT is outperformed by Duluth-type treatment. While some may attempt to selectively use these data to bolster their arguments, claims for the superiority of one treatment type over another is unwarranted.

In retrospect, it is not surprising that there were no significant differences between CBT and Duluth-type interventions. Modern batterer groups tend to mix different theoretical approaches to treatment, combining both feminist theory of power and control as well as specific interventions to deal with anger control, stress management, and improved communication skills (Davis & Taylor, 1999; Healy et al., 1998). The "brand name" labels can be misleading. No researchers to date have conducted a head-to-head comparison between CBT and Duluth-type battering interventions, perhaps due the difficulty in identifying treatment techniques unique to either school.

It is common in the psychotherapy outcome literature to find that different modalities of treatment are equally effective—and to conclude that all have won (Beutler, 1991). This phenomenon of finding comparability in treatment outcomes is referred to as the "dodo bird verdict" (Beutler, 1991; Luborsky et al., 1975). Equivalent effect sizes due to treatment are common results of comparative studies of two active treatments (DeRubeis & Crits-Cristoph, 1998). In this case, only one study has conducted a randomized clinical trial of two active treatments (CBT and couples groups) against a no-treatment control (Dunford, 2000). Within this study and across the domain of studies to date, effects sizes due to all types of interventions are small.

### 4.3. *Have all lost?*

While the effect size due to treatment overall is in the small range, there are some specific studies finding large effect sizes. As shown in [Table 1](#), the interventions with the largest effect sizes were obtained from 16-week group therapies supplemented with retention techniques ([Taft et al., 2001](#)) and 12-week relationship enhancement skills training groups ([Waldo, 1988](#)). These findings can either be dismissed as “outliers” among scientific treatment studies, or viewed as harbingers of potentially powerful interventions. In the first study, [Taft et al. \(2001\)](#) randomly assigned men to either CBT or supportive therapy groups, both of which were supplemented with techniques designed to improve treatment retention based on the principles of motivational interviewing ([Miller & Rollnick, 1991](#)). These techniques consisted of reminder phone calls and supportive handwritten notes after intake and after missed sessions. As a result, the authors report one of the lowest attrition rates in the literature. The core therapies differed dramatically from one another, one being highly structured and the other unstructured, but both revealed strong effect sizes, especially when based on police report. This study suggests that the small effect sizes due to batterers’ interventions may be in part attributable to the client’s noninvestment and subsequent attrition from the programs. These simple techniques, which can be an adjunct to any type of program, may increase the client’s perception that the program is aware of his absence and is invested in his welfare. Thus, he may be more motivated to complete and actively participate in the program, lowering attrition and recidivism.

The second study to find a large effect size was an evaluation of an intervention called relationship enhancement ([Guernsey, 1977](#)). The goals of relationship enhancement as applied to battering are to help the men develop interpersonal skills that enhance relationships and enable them to stop their use of violence ([Waldo, 1988](#)). Interventions include role-plays and assigned homework targeted to improve expressive skills, empathy, communication with the partner, and the identification and management of their emotions (see [Waldo, 1985](#)). This study suggests that more emotion-focused, rather than cognitively focused, interventions may increase the effect size of batterers treatment. Of course, the results of any single, unreplicated study should not be over generalized. More research is needed on the effectiveness of motivational interviewing as well as emotion-focused approaches as treatment modalities or as additive components to existing batterers’ intervention groups.

### 4.4. *Limitations*

One of the greatest concerns when conducting a meta-analysis is the ease at which the “bottomline” is recalled and the extensive caveats for caution are forgotten or ignored. Although we selected only studies that met our minimal criteria for rigor (inclusion of a comparison group, a follow-up period beyond the end of treatment, not relying on batterers’ self-report), there remains significant variability in the quality of research studies. Even the experimental studies are hindered by problems with high attrition rates, inconsistencies in reporting recidivism for dropouts, and low reporting rates at follow-up ([Gondolf, 2001](#)). Some of these factors that affect the quality of the research studies are confounded with



treatment quality and quality of the community response, broader factors that cannot always be ascertained. Therefore, caution in interpreting these results is warranted. Meta-analyses are only as robust as the individual studies taken into account.

Quasi-experiments make up the bulk of the studies included in this meta-analysis, but studies comparing treatment completers to dropouts are inherently confounded by self-selection. Quasi-experiments capitalize on “creaming” (Davis & Taylor, 1999); that is, comparing the most highly motivated batterers with the least motivated batterers, “thereby stacking the deck in favor of finding program effects” (Davis & Taylor, 1999, p. 74). Yet, experiments have interpretational biases as well. Most studies base outcomes according to the original random assignment. If the experimental treatment suffers from high attrition rates, and the outcome data are based on “intention to treat,” there is a strong possibility that few people received an adequate “dose” of treatment (Gondolf, 2001). The alternative, using treatment actually received, results in a violation of random assignment while simultaneously engaging in “creaming,” making the experiment no more rigorous than a quasi-experiment. Policymakers want to know whether mandating counseling leads to lower rates of recidivism in comparison to other approaches. This question has two parts: (1) Will they attend treatment if mandated? (2) Will treatment have an impact on recidivism if they do attend? Both true and quasi-experiments must grapple with how to tease apart the two parts of this question.

Other limitations include variability across studies concerning what constitutes successful treatment completion. In some cases, the definition was clear (e.g., completing 70% or 80% of the required sessions) and in other studies, it was unspecified. Future researchers should carefully specify what qualifies as successful completion of treatment and also examine the relationship between number of treatment sessions attended and outcome to identify any potential “dose–response” curve. The reliance on dichotomous variables of recidivism may be an overly conservative estimate and dampen the effect size of batterers’ treatment. The overall effect sizes may be larger if one uses a reduction of violence rather than cessation of violence as the outcome measure. However, doing so would result in the inclusion of a smaller number of studies, as several early studies do not report the necessary statistics. In addition, the clinical significance of the change in violence attributable to batterers’ intervention may be questionable.

All longitudinal studies are affected by follow-up rates. As shown in Tables 1 and 2, many studies fail to report participation rates of partners at follow-up. Where partner follow-up contact rates are reported, they range from 22% to 90% of the sample. Those who are lost to follow-up are thought to be more abusive (DeMaris & Jackson, 1987), and therefore success rates may be inflated (Tolman & Bennett, 1990). As such, the resultant effect sizes would also probably be overestimates.

Like partner reports, police reports as outcome measures of recidivism are also problematic and may not adequately reflect reality. With couples already involved in family violence interventions, only about one out of every five domestic violence assaults are reported to the authorities (Rosenfeld, 1992). In some jurisdictions, police reports themselves are inaccurate. Crimes committed outside of the state or local jurisdiction, or incidents of violence in which adjudication was deferred may not appear on the criminal record. Crimes that do appear on the record may be ambiguous as to whether they were family violence or other types of

assault, and researchers have to grapple with which types of crimes “count” in terms of recidivism.

Moreover, the effect size due to treatment for court-mandated batterers is confounded with the strength of the coordinated efforts of the police, probation, and legal system. The potency of the legal system that sanctions men for noncompliance may have a profound effect on treatment completion rates and, as a result, the effect of treatment. Yet, few studies attempt to examine the additive effects of arrest, prosecution, treatment, probation, and legal action for noncompliance (Babcock & Steiner, 1999; Murphy, Musser, & Maton, 1998, are exceptions).

Given these methodological and pragmatic issues, it is not surprising that the effect sizes attributable to batterers’ treatment are small. Although we excluded treatment comparison studies because they only allow an estimate of the size of the difference between two active interventions, the entire literature on batterers’ intervention is actually predominated by component analysis studies, attempting to measure the additive component of the treatment on top of the legal interventions. Since involvement in the legal system is probably beneficial in reducing recidivism (Dutton, 1987), court-ordered treatment programs must reduce abuse recidivism further to demonstrate the effectiveness of treatment over and above legal-system interventions (Rosenfeld, 1992). Differences between two active interventions are more difficult to find than between treatment and no-treatment conditions. Added to that is the spontaneous violence cessation rate in nonclinical samples of about 35% (O’Leary et al., 1989). For batterers’ interventions to be proven effective, they must supercede both the spontaneous recovery rate and the effects of legal interventions.

#### *4.5. Clinical and policy implications*

Policymakers should not accept the null hypothesis and dismiss the potential for batterers’ interventions to have an impact on intimate partner abuse. Results showing a small effect of treatment on violence abstinence do not imply that we should abandon our current battering intervention programs. Similar small treatment effects are found in meta-analyses of substance abuse treatments when abstinence from alcohol is the outcome of interest (Agosti, 1995). Yet, some people are able to dramatically transform their lives following substance abuse or battering interventions. Given what we now know about the overall small effect size of batterers’ treatment, the energies of treatment providers, advocates, and researchers alike may best be directed at ways to improve batterers’ treatment. Because no one treatment model or modality has demonstrated superiority over then others, it is premature for states to issue mandates limiting the range of treatment options for batterers. Battering intervention agencies are more likely to improve their services by adding components or tailoring their treatments to specific clientele, than by rigidly adhering to any one curriculum in the absence of empirical evidence of its superior efficacy. Different types of batterers may preferentially benefit from specific forms of interventions (Saunders, 1996), yet no controlled treatment-matching studies have been conducted to date. While a small number of studies have assessed group and couples’ formats, no published studies to date have attempted to assess the efficacy of individual treatment for battering, although this researchers are embarking on this frontier (e.g., Fruzzetti, 2001; Rathus, 2001). Promising directions for improving treatment efficacy

include targeting treatments to specific subsamples, such as different ethnic minority groups, batterers who are chemically dependent, batterers at different motivational stages, different types of batterers (e.g., family-only, borderline, and antisocial/generally violent types), and women arrested for domestic violence. Treatment providers should develop alternative techniques and collaborate with researchers to evaluate their efficacy in an effort to develop evidence-based practice. To this end, researchers need to become an integral part of the coordinated community response to domestic violence.

Batterers' treatment is just one component of the coordinated community response to domestic violence. Police response, prosecution, probation, as well as treatment all affect recidivism of domestically violent partners. Even the best court-mandated treatment programs are likely to be ineffective in the absence of a strong legal response in initial sentencing and in sanctioning offenders who fail to comply with treatment. Even then, treatment may not be the best intervention for all batterers. Alternative sanctions should be developed and empirically tested along with alternative treatments.

## References

References marked with an asterisk indicate studies included in the meta-analysis.

- Abel, E. (in press). Treatment for female "batterers." *Journal of Family Violence*.
- Adams, D. A. (1988). Counseling men who batter: A pro-feminist analysis of five treatment models. In M. Bograd, & K. Yllo (Eds.), *Feminist perspectives on wife abuse* (pp. 176–199). Newbury Park, CA: Sage.
- Agosti, V. (1995). The efficacy of treatments in reducing alcohol consumption: A meta-analysis. *International Journal of the Addictions*, 30, 1067–1077.
- Austin, J. B., & Dankwort, J. (1999). Standards for batterer programs: A review and analysis. *Journal of Interpersonal Violence*, 14, 152–168.
- Babcock, J. C., & LaTaillade, 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., & Steiner, R. (1999). The relationship between treatment, incarceration, and recidivism of battering: A program evaluation of Seattle's coordinated community response to domestic violence. *Journal of Family Psychology*, 13, 46–59.
- Beutler, L. (1991). Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. *Journal of Consulting and Clinical Psychology*, 59, 226–232.
- Bargarozzi, D. A., & Giddings, C. W. (1983). Conjugal violence: A critical review of current research and clinical practices. *The American Journal of Family Therapy*, 11, 3–15.
- 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, 405–416.
- Bureau of Justice Statistics (1998). *Violence by intimates: Analysis of data on crimes by current or former spouses, boyfriends, and girlfriends*. Washington, DC: U.S. Department of Justice, NCJ-167237.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Co.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Cristoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *Clinical Psychologist*, 49, 5–11.
- \*Chen, H., Bersani, S., Myers, S. C., & Denton, T. (1989). Evaluating the effectiveness of a court-sponsored abuser treatment program. *Journal of Family Violence*, 4, 309–322.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Davis, R. C., & Taylor, B. G. (1999). Does batterer treatment reduce violence? A synthesis of the literature. *Women and Criminal Justice, 10*, 69–93.
- \*Davis, R. C., Taylor, B. G., & Maxwell, C. D. (1998). Does batterer treatment reduce violence? A randomized experiment in Brooklyn. *Justice Quarterly, 18*, 171–201.
- DeMaris, A., & Jackson, J. K. (1987). Batterers reports of recidivism after counseling. *Social Casework, 68*, 458–465.
- DeRubeis & Crits-Cristoph (1998). Empirically supported individual and group psychological treatment for adult mental disorders. *Journal of Consulting and Clinical Psychology, 66*, 37–52.
- \*Dobash, R., Dobash, R. E., Cavanagh, K., & Lewis, R. (1996). Reeducation programs for violent men—An evaluation. *Research Findings, 46*, 309–322.
- Dunford, F. W. (1998, July). *Experimental design and program evaluation*. Paper presentation at the Program Evaluation and Family Violence Research: An International Conference, Durham, NH.
- \*Dunford, F. W. (2000). The San Diego Navy experiment: An assessment of interventions for men who assault their wives. *Journal of Consulting and Clinical Psychology, 68*, 468–476.
- Dutton, D. G. (1987). The criminal justice system response to wife assault. *Law and Human Behavior, 11*, 189–206.
- \*Dutton, D. G., Bodnarchuk, M., Kropp, R., & Hart, S. D. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. *International Journal of Offender Therapy and Comparative Criminology, 41*, 9–23.
- Dutton, D. G. (1998). *The abusive personality: Violence and control in intimate relationships*. New York: Guilford Press.
- Edleson, J. (1996). Controversy and change in batterers' programs. In J. L. Edleson, & Z. C. Eisikovits (Eds.), *Future interventions with battered women and their families* (pp. 154–169). Thousand Oaks, CA: Sage.
- Edleson, J., & Grusznski, R. (1988). Treating men who batter: Four years of outcome data from the domestic abuse project. *Journal of Social Service Research, 12*, 3–22.
- \*Edleson, J. L., & Syers, M. (1991). The effects of group treatment for men who batter: An 18-month follow-up study. *Research in Social Work Practice, 1*, 227–243.
- Fagan, J. (1989). Cessation of family violence: Deterrence and dissuasion. In L. Ohlin, & M. Tonry (Eds.), *Family Violence: Crime and Justice: A review of research* (pp. 377–425). Chicago: The University of Chicago.
- \*Feder, L., & Forde, D. (1999, July). *A test of the efficacy of court-mandated counseling for convicted misdemeanor domestic violence offenders: Results from the Broward experiment*. Paper presented at the International Family Violence Research Conference, Durham, NH.
- \*Flourmoy, P.S. (1993). A comparison of groups for men who batter (Doctoral dissertation, Washington State University), *Dissertation Abstracts International, 53*, 5974.
- \*Ford, D. A., & Regoli, M. J. (1993). The criminal prosecution of wife batterers: Process, problems, and effects. In N. Z. Hilton (Ed.), *Legal responses to wife assault* (pp. 127–164). Newbury Park, CA: Sage.
- Fruzzetti, A. E. (2001, November). Dialectical Behavior Therapy adapted for treatment of partner-violent men. *Panel discussion at the Association for the Advancement of Behavior Therapy Convention* (Philadelphia, PA.).
- Gelles, R. J. (1980). Violence in the family: A review of research in the seventies. *Journal of Marriage and the Family, 42*, 873–885.
- \*Gondolf, E. (1998). Do batterer programs work? A 15-month follow-up of a multi-site evaluation. *Domestic Violence Report, 3*, 64–65, 78–79.
- Gondolf, E. (2000). Reassault at 30-months after batterer programs intake. *International Journal of Offender Therapy and Comparative Criminology, 44*, 111–128.
- Gondolf, E. (2001). Limitation of experimental evaluations of batterer programs. *Trauma, Violence, and Abuse, 2*, 79–88.
- Gondolf, E. W. (1997). Patterns of reassault in batterer programs. *Violence and Victims, 12*, 373–387.
- Guernsey, B. G. (1977). *Relationship enhancement: Skill training programs for therapy, problem prevention, and enrichment*. San Francisco: Jossey-Bass.

- \*Hamberger, K., & Hastings, J. (1988). Skills training for treatment of spouse abusers: An outcome study. *Journal of Family Violence*, 3, 121–130.
- Hamberger, K., & Hastings, J. (1993). Court-mandated treatment of men who assault their partner: Issues, controversies, and outcomes. In N. Z. Hilton (Ed.), *Legal responses to wife assault* (pp. 188–229). Newbury Park, CA: Sage.
- \*Harrell, A. (1991). *Evaluation of court ordered treatment for domestic violence offenders* (Final Report). Washington, DC: Urban Institute.
- \*Hawkins, R., & Beauvais, C. (1985, August). *Evaluation of group therapy with abusive men: The police record*. Paper presented at the meeting of the American psychological association, Los Angeles, CA.
- Healey, K., Smith, C., O'Sullivan, C. (1998). *Batterer intervention: Program approaches and criminal justice strategies*. Report to the National Institute of Justice, Washington, DC.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods*, 1, 154–169.
- Holtzworth-Munroe, A., Beatty, S. B., & Anglin, K. (1995). The assessment and treatment of marital violence: An introduction for the marital therapist. In N. S. Jacobson, & A. S. Gurman (Eds.), *Clinical handbook of couple therapy* (pp. 317–339). New York: Guilford Press.
- Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 80, 327–334.
- Jacobson, N. S. (1993, October). *Domestic violence: What are the marriages like?* Plenary delivered at the annual meeting of American Association of Marriage and Family Therapy, Anaheim, CA.
- Johnson, B. T. (1995). *D-Stat: Software for the meta-analytic review of research literatures* (Version 1.11). Hillsdale, NJ: Erlbaum.
- \*Leong, D. J., Coates, C. J., & Hoskins, J. (1987). *Follow-up of batterers treated in a court-ordered treatment program*. Paper presented at the Third National Family Violence Research Conference, University of New Hampshire, Durham, NH.
- Levesque, D. A., & Gelles, R. J. (1998, July). *Does treatment reduce recidivism in men who batter? A meta-analytic evaluation of treatment outcome*. Paper presentation at the Program Evaluation and Family Violence Research: An International Conference, Durham, NH.
- Lipchik, E., Sirles, E. A., & Kubicki, A. D. (1997). Multifaceted approaches in spouse abuse treatment. In R. Geffner, S. B. Sorenson, & P. K. Lundberg-Love (Eds.), *Violence and sexual abuse at home: Current issues in spousal battering and child maltreatment* (pp. 131–148). New York: Haworth Press.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181–1209.
- Loesel, F., & Koeferl, P. (1987). Evaluationsforschung zur sozialtherapeutischen Anstalt: Eine meta-analyse (Evaluation research on the social-therapeutic prison: A meta-analysis). *Gruppendynamik*, 18, 385–406.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that “everyone has won and all must have prizes”? *Archives of General Psychiatry*, 32, 995–1008.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, 71, 173–180.
- Miller, W. R., & Rollnick, S. (1991). *Motivational interviewing: Preparing people to change addictive behavior*. New York: Guilford Press.
- Moffit, T. E., Caspi, A., Krueger, R. F., Magdol, L., Margolin, G., Silva, P. A., & Sydney, R. (1997). Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Psychological Assessment*, 9, 47–56.
- \*Morrel, T. M., Elliott, J. D., Murphy, C. M., & Taft, C. (2003). A comparison of cognitive-behavioral and supportive group therapies for male perpetrators of domestic abuse. *Behavior Therapy*, 24, 77–95.
- Murphy, C. M., & Baxter, V. A. (1997). Motivating batterers to change in the treatment context. *Journal of Interpersonal Violence*, 12, 607–619.

- \*Murphy, C. M., Musser, P. H., & Maton, K. I. (1998). Coordinated community intervention for domestic abusers: Intervention system involvement and criminal recidivism. *Journal of Family Violence, 13*, 263–284.
- \*Newell, R. G. (1994). The effectiveness of court-mandated counseling for domestic violence: An outcome study. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 53*, 1193.
- O'Leary, K. D., Barling, J., Aria, I., Rosenbaum, A., Malone, J., & Tyree, A. (1989). Prevalence and stability of physical aggression between spouses. A longitudinal analysis. *Journal of Consulting and Clinical Psychology, 57*, 263–268.
- O'Leary, K. D., Heyman, R. E., & Neidig, P. H. (1999). Treatment of wife abuse: A comparison of gender-specific and conjoint approaches. *Behavior Therapy, 30*, 475–506.
- \*Palmer, S. E., Brown, R. A., & Barrera, M. E. (1992). Group treatment program for abusive husbands: Long term evaluation. *American Journal of Orthopsychiatry, 62*, 276–283.
- Peluch, J. (1987). *Goals for men's treatment groups*. Unpublished manuscript, Family Services of Hamilton-Wentworth, Ontario, Canada.
- Pence, E., & Paymar, M. (1993). *Education groups for men who batter: The Duluth model*. New York: Springer-Verlag.
- Rathus, J. H. (2001, November). *Dialectic behavior therapy adapted for the treatment of partner violent men*. Panel presentation at the Association for the Advancement of Behavior Therapy, Philadelphia, PA.
- Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (1997). *Metawin: Statistical software for meta-analysis with resampling tests*. Sunderland, MA: Sinauer Associates.
- Rosenfeld, B. D. (1992). Court ordered treatment of spouse abuse. *Clinical Psychology Review, 12*, 205–226.
- Rosenthal, R. (1995). Writing meta analytic reviews. *Psychological Bulletin, 118*, 183–192.
- Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology, 35*, 203–208.
- Saunders, D. G. (1996). Feminist–cognitive–behavioral and process–psychodynamic treatments for men who batter: Interaction of abuser traits and treatment model. *Violence and Victims, 11*, 393–414.
- Shadish, W. R., & Ragsdale, K. (1996). Random versus nonrandom assignment in controlled experiments: Do you get the same answer? *Journal of Consulting and Clinical Psychology, 64*, 1290–1305.
- Sherman, L. W., Smith, D., Schmidt, J. D., & Rogan, D. P. (1992). Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review, 57*, 680–690.
- Smith, M. L., Glass, G. V., & Miller, T. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Sonkin, D. J., Martin, D., & Walker, L. E. A. (1985). *The male batterer: A treatment approach*. New York: Springer-Verlag.
- \*Stacey, W. A., & Shupe, A. (1984). *An evaluation of three programs for abusive men in Texas* (Research Monograph No. 29). Arlington, TX: Center for Social Research, University of Texas.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41*, 75–88.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scales (CTS-2). *Journal of Family Issues, 17*, 283–316.
- Sullivan, C. M., & Bybee, D. I. (1999). Reducing violence using community based advocacy for women with abusive partners. *Journal of Consulting and Clinical Psychology, 67*, 43–53.
- \*Taft, C. T., Murphy, C. M., Elliot, J. D., & Morrel, T. M. (1990). Attendance enhancing procedures in group counseling for domestic abusers. *Journal of Counseling Psychology*.
- Tolman, R. T., & Bennett, L. (1990). A review of quantitative research on men who batter. *Journal of Interpersonal Violence, 5*, 87–118.
- U.S. Attorney General (1994). *Task Force on Family Violence: Final report*. U.S. Department of Justice.
- Waldo, M. (1985). A curative factor framework for conceptualizing group counseling. *Journal of Counseling and Development, 64*, 52–58.
- \*Waldo, M. (1988). Relationship enhancement counseling groups for wife abusers. *Journal of Mental Health Counseling, 10*, 37–45.

- Wang, M. C., & Bushman, B. J. (1998). Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods*, 3, 46–54.
- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117, 450–468.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage.
- Zimring, F. E. (1989). Toward a jurisprudence of family violence. In L. Ohlin, & M. Tonry (Eds.), *Family violence: Crime and justice: A review of research* (pp. 547–569). Chicago: University of Chicago Press.