



Pergamon

Aggression and Violent Behavior
9 (2004) 605–631

AGGRESSION
AND
VIOLENT
BEHAVIOR

Evaluating batterer counseling programs: A difficult task showing some effects and implications

Edward W. Gondolf*

*Mid-Atlantic Addiction Training Institute (MAATI), Indiana University of Pennsylvania (IUP),
Indiana, PA 15705, USA*

Received 19 February 2003; received in revised form 6 June 2003; accepted 19 June 2003

Abstract

Over 40 published program evaluations have attempted to address the effectiveness of “batterer programs” in preventing reassaults. Summaries and meta-analysis of these evaluations suggest little or no “program effect.” Methodological shortcomings, however, compromise most of these quasi-experimental evaluations. Three recent experimental studies appear to confirm little or no effect, but implementation problems, intention-to-treat design, and sample attrition limit these results. A longitudinal 4-year follow-up evaluation in four cities poses additional considerations and evidence of at least a moderate program effect. There is a clear deescalation of reassault and other abuse, the vast majority of men do reach sustained nonviolence, and about 20% continuously reassault. The prevailing cognitive-behavioral approach appears appropriate for most of the men, but the following enhancements are warranted: swift and certain court response for violations, intensive programming for high-risk men, and ongoing monitoring of risk. Program effectiveness depends substantially on the intervention system of which the program is a part.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Program evaluation; Batterer counseling; Domestic violence; Batterers

* Tel.: +1-724-357-4749; fax: +1-724-357-3944.

E-mail address: egondolf@iup.edu (E.W. Gondolf).

URL: <http://www.iup.edu/maati/publications>.

1. Introduction

1.1. Emergence of batterer programs

The question of what to do about men who batter their female partners has haunted the domestic violence field since its emergence in the late 1970s. Many advocates working with battered women felt—and many still feel—that few batterers could be changed given the social reinforcement for and tolerance of violence against women (Taubman, 1986). Trying to counsel or educate such men might, in fact, raise false hopes in battered women and worsen their already difficult circumstances. Protection for women and separation from their male batterers, therefore, became the overarching intervention objective (e.g., Dobash & Dobash, 1992). In the late 1970s, however, some men's counselors allied with the battered women's advocates and began antisexist consciousness-raising groups primarily for men who professed wanting to change (Adams, 1988). Group facilitators led discussions that exposed men's socialization to dominate women and, in some cases, use violence to maintain that dominance (see Adams, 1989).

Batterer programs gradually became more sophisticated by adopting cognitive–behavioral techniques from the counseling of other violent men (Pence & Paymar, 1993). Some of these programs emphasizing gender issues have been accused of being too “confrontive” (Stosny, 1995), while others emphasizing skill building are often criticized for being naïvely superficial and lacking a clear message of change (Gondolf & Russell, 1986). The fact is that numerous curriculums have been developed incorporating both gender issues and cognitive–behavioral techniques (e.g., Kivel, 1992; Pence & Paymar, 1993; Russell, 1995; Stordeur & Stille, 1989).

While a gender-based cognitive–behavioral orientation is the most prominent approach, other more psychodynamic or emotive approaches have been forwarded as well (e.g., Dutton, 1998; Stosny, 1995). These approaches attempt to address the psychological issues and emotional hurts of men that may contribute to their abuse. A variety of other therapies have also emerged: anger management (a streamlined adaptation of cognitive–behavioral treatment), dialectical counseling, neuropsychological treatment, “wrap-around” services, and couples counseling (for a review of couples counseling, see O'Leary, 2002).

1.2. Gender-based, cognitive–behavioral convergence

Despite caricatures of some conflicting approaches (e.g., Raab, 2000), there has been a convergence around the fundamental parameters of gender-based, cognitive–behavioral counseling (Austin & Dankwort, 1999; Bennett, 1998). This approach generally prompts an individual to recognize and take responsibility for his abusive behavior, use techniques to avoid abuse, develop alternative behaviors, and expose rationalizations for abuse. Of course, within these parameters are a range of styles, experience, training, and orthodoxy. There is also a range of formats from educational or instructional formats to those that are more discussion- and process-oriented. One of the main questions from this 20-year evolution is whether or not the resultant gender-based, cognitive–behavioral counseling is effective in

reducing men's abuse, and particularly their reassault, of women. Reassault has been the principal outcome of interest since it is associated with physical injury, is the prime concern of the courts, and is more concretely measurable.

The following discussion explores the attempts to answer the pressing question of program effectiveness, and in the process shows why an answer has been difficult to reach. It begins by illustrating the conceptual and methodological challenges in evaluating the effectiveness of batterer programs and reasons for some of the contradictory outcomes. There are substantial conceptual and methodological issues that need to be considered in implementing and interpreting especially the recent experimental program evaluations. These issues, and the small number of substantial evaluations, undercut the utility of the recent meta-analyses in the field. The meta-analyses also neglect the findings of longitudinal trends, sophisticated dose–response analysis, and the circumstantial evidence from complementary research.

Next, we summarize various findings from our 7-year multisite evaluation of batterer programs that attempts to address some of these issues and shortcomings. We construct a case for program effectiveness and improvement by weighing various outcomes from our multisite evaluation with complementary studies. Our conclusion is that some batterer counseling programs appear to contribute to the reduction of reassault, but this possibility is related to the intervention system of which the program is a part. The majority of the men, regardless of “personality type,” appear to benefit from the intervention. Categorical claims that batterer programs “do not work” may, therefore, over generalize or misinterpret the current research.

Our review of the literature constructs a case for program effectiveness through synthesizing a patchwork of research that appears to intersect and even converge at points around the strengths of a comprehensive multisite evaluation. The result is more an overview of critical issues and considerations in batterer program evaluation rather than a systematic dissection of evaluations or technical account of our multisite evaluation. Several research reviews of this kind have already been published by both practitioner–researchers (e.g., Aldarondo, 2002; Mederos, 1999) and researcher–practitioners (e.g., Babcock & LaTaillade, 2000; Saunders & Hamill, *in press*), and over 40 technical research articles have been published on our multisite evaluation.

2. Evaluation issues

2.1. Program definitions

Evaluating the effectiveness of batterer programs is a difficult and complex task that complicates the interpretation of the evaluation results. As has been the case in other fields, such as alcohol, sex offense, and depression treatment, different program conceptions, outcome measures, research designs, and statistical analyses can produce contrary results (e.g., Elkin, Shea, & Watkins, 1989; National Opinion Research Center, 1996; Project MATCH Research Group, 1997). Moreover, similar programs with different men or in different settings can have different outcomes. Program evaluations that specifically address

these sorts of issues are likely to further their validity, and those that at least acknowledge them will help clarify interpretation of the results.

The first major set of issues is those that deal with program definition. These issues are probably the least discussed and refined in the evaluations of batterer programs thus far. Is the “program” merely the group counseling within a weekly, 1.5- to 2-h session? Is it also the program intake, individual assessment or screening, dismissal policies, and reporting procedures? Does it include the outreach to and support of the victims that some programs offer? The evolving nature of especially community-based programs complicates the answer to these questions. Protocols, curriculums, counselor skill, administration, and referral and reporting procedures tend to change and develop. Does the evaluation outcome at the end of an extended follow-up apply to the current program that may have evolved from the initial one? Our multisite study documented substantial changes at the program sites following our period of subject recruitment and treatment (see [Gondolf, 2002a](#)). New program directors are currently at three of the four sites, half to all of the staff have been replaced, one program has an entirely new curriculum, the two 3-month programs have been extended to 4 and 6 months, and the court referral system at two of the sites has been substantially revised.

Additionally, batterer programs are enmeshed in an elaborate intervention system that includes police practices, court action, probation supervision, civil protection orders, victim services, additional services for the men, community resources, and local norms (see [Gamache, Edleson, & Schock, 1988](#); [Murphy, Musser, & Maton, 1998](#)). These components ideally work together within what has come to be called a “community coordinated response” ([Pence & Shepard, 1999](#)). This sort of system varies considerably in extent, development, and actual coordination under so-called domestic violence councils. The councils also run the gambut in participation, viability, and authority. To what extent do we separate and distinguish batterer counseling from these components, and to what degree do the components contribute to the counseling? Some observers, in fact, argue that such components cannot be separated since they combine in a synergetic effect toward a “tipping point” of change ([Gladwell, 2000](#)). In summary, batterer programs are part of a dynamic context that needs to be weighed in analyzing and interpreting outcomes.

2.2. Measuring outcomes

The most obvious set of issues is around the measurement of outcome: What, how, and when to assess the outcome? While reassault has been the most frequent “what” used as an outcome, there are varying measurement tools, data collection techniques, and circumstances to consider. Do we tally items on inventories, such as the Conflict Tactics Scale ([Straus, 1979](#)), into an interval score, or do we classify the items into categories of abuse? Most evaluations have used a fairly dichotomous outcome, such as “reassault” versus “no reassault,” but multiple outcomes, which include different levels and patterns of abuse, are the ideal. There is also concern that assault is only the “tip of the iceberg.” A constellation of abuses that includes controlling behavior, verbal abuse, and threats, and ultimately the woman’s general well-being or quality-of-life needs to be considered ([Dobash, Dobash, Cavanagh, & Lewis, 2000](#); [Smith & Smith, 1999](#)).

The outcome period is also a crucial issue. What is considered a sufficient duration to test—6 months, 1 year, 2 years? When does the period start—at program intake or after the program when the men have received a sufficient “dose” of counseling? If we have programs of different lengths, should we adjust the outcome periods so that they have equivalent postprogram follow-ups? One of the answers emerging from the evaluations in other fields is the shift from cumulative outcomes to longitudinal retrospective ones (i.e., how many months has the program participant been violent-free *previous* to the follow-up endpoint). For example, “recovery” from alcohol misuse appears to be nonlinear with points of relapse. The current trend in alcohol treatment evaluation is, therefore, to measure periods or days of sobriety or current alcohol use rather than the cumulative rate of relapse (e.g., [National Opinion Research Center, 1996](#)). These different approaches to measuring outcome can shift apparent “failure” to “success,” as we will illustrate later with our multisite evaluation.

Finally, we are faced with a fundamental issue of interpretation: How much of an effect constitutes success or warrants an endorsement (see [Jennings, 1990](#))? What do we mean by “effective” or that a program “works.” Program evaluators have questioned: Effective compared with what, with whom, and under what circumstances? There is the additional question being raised in these days of “managed care:” Is the effect worth the cost to insurance companies or state agencies? In the case of batterer programs: Is the cost diverting needed and deserved funding from victim services, or is it helping the victims through beneficial outcomes? The answers to these questions are largely subjective and require an interpretation of the outcome using the nuance of qualitative research and clinical experience, as well as a cost analysis (see [Jones, 2000](#)).

2.3. *Research design*

2.3.1. *Experimental designs*

The biggest and most controversial issue surrounding program evaluation is whether to use experimental or quasi-experimental designs. While the former is considered more “scientific,” experimental designs encounter substantial implementation challenges and conceptual shortfalls. Some evaluators question whether they can truly be achieved in criminal justice settings and end up causing more disruption and misinformation than they are worth ([Marshall & Serran, 2000](#)). But can quasi-experimental and other alternative designs meaningfully address program effectiveness without an equivalent control group? As illustrated below, our experience is that current analytical procedures make it possible to simulate experimental conditions within quasi-experimental evaluations and to do so in a way that represents a “dose response” to counseling.

Most researchers in the positivist tradition hold experimental designs to be the “gold standard” for evaluation ([Boruch, Snyder, & DeMoya, 2000](#); [Dunford, 2000a](#)). Randomly assigning subjects to an experimental (i.e., treatment) group and a control (i.e., nontreatment) group, as a means to maximize internal validity, seems straightforward enough. Unfortunately, implementation of such designs is especially problematic, as an extensive literature in the medical and public health field shows. But experimental evaluations raise conceptual issues, as well. The primary one has to do with the “intention to treat” approach assumed in

most experimental designs, as opposed to a dose response (Efron & Feldman, 1991). In experimental designs, the experimental group consists of everyone sent to the program or treatment, whether they receive the treatment or dropout. The drop outs—or those with a “low dose” of the program—can cancel out the apparent effectiveness of the program completers—or “high dose” participants. The comparison of an experimental group versus control group, therefore, may tell less about treatment effectiveness and more about the procedures of referring to and retaining men in a certain program.

The effectiveness of the referral may be as much dependent on the intervention system as a whole as it is on the program. Some programs screen out or dismiss men who do not appear “motivated” to change; others attempt to motivate men in program orientations or through the threat of sanctions. However, most batterer programs monitor the men’s attendance and return them to the court system if they do not attend. These programs contend that the responsibility to keep the men in the program is not theirs but rather the responsibility of the courts that refer the men. There are analytical steps to assess the dose response in an experimental design, but they essentially turn the design into a quasi-experiment and require large samples and extensive intake assessment to accomplish.

Another major implementation issue, especially in the courts, is finding a pure control group. Men assigned to open probation are generally used as the control group, but many of these men, or their partners, receive compensating treatments, sanctions, or surveillance in a viable “coordinated community response.” For instance, we found women whose partners dropped out early were more likely to seek additional help from a women’s center. Moreover, an experiment can be disruptive to the intervention system and actually change it. The results may end up reflecting an altered system rather than the initially intended one. This issue has especially been evident in recent experimental evaluations of batterer programs (e.g., Feder & Dugan, 2002; Feder, Jolin, & Feyerharm, 2000).

2.3.2. *Quasi-experimental designs*

Most evaluations of batterer programs have loosely followed a quasi-experimental design, mostly for practical reasons rather than conceptual ones. Quasi-experimental designs generally compare those completing the program (or receiving a certain dose in terms of attendance) to a group of program dropouts or “no-shows.” The proverbial criticism is that the comparison ends up one of “apples and oranges” since dropouts generally have different characteristics than the program completers (see Daly, Powers, & Gondolf, 2001; Rondeau, Brodeur, Brochu, & Lemire, 2001). The notion of dropout is, furthermore, a poorly defined one (Daly & Pelowski, 2000). Dropouts can be those who are dismissed by a program for absences or failure to pay, or those who willingly withdraw or reassault their partners. These policies and practices of a program regarding dropout—which vary substantially across and within programs—influence who ends up in the “quasi-control” group of dropouts. The main attraction of quasi-control groups is that they are easy to implement, consider programs in their natural state, and are based on a dose response.

There are analytical procedures to address the limitations of quasi-experimental designs. The most common is to control for potential differences between the dropout and completer groups with a logistic regression. This procedure requires a comprehensive set of intake or

control variables and a large sample of respondents. But a more fundamental problem leads statisticians to consider logistic regression as a “naïve” analysis. The regression treats program dropout or attendance as an independent variable when it is “endogenous” or dependent on many of the same controls that influence reassault. Statisticians in the public health field have long faced this problem since so many community-based interventions do not readily lend themselves to experimental evaluation (e.g., how would you test to see if preventing teen pregnancy improves adult physical health). One popular option has been the use of Propensity Score analysis which develops a weighting for the likelihood to drop out and uses that weighting, or “score,” to match subjects in the dropout and completer groups (Jones, D’Agostino, Gondolf, & Heckert, *in press*). (A variation of Propensity Score analysis directly adjusts the effect of program completion on reassault.)

2.3.3. *Alternative designs*

An alternative view is that both pure and quasi-experiments are not sufficiently “realistic” (Dobash & Dobash, 2000; Pawson & Tilly, 1997; Van Voohis, Cullen, & Applegate, 1995). That is, they often manipulate the intervention for random assignment or impose a prescribed treatment. But most of all, they fail to account for the context of the program that may substantially contribute to the programs outcomes. This context includes not only the immediate context of the intervention system components but also additional interventions, help-seeking, and circumstances (e.g., drinking, unemployment) during the follow-up. A more realistic outcome might be, therefore, a dynamic conception that includes mediating individual conditional factors and community resources and supports (see Mulvey & Lidz, 1985; Steadman, 1982). The outcomes may be influenced, for example, by women’s counseling or men’s alcohol treatment during the follow-up or by the police practices and service availability in the community.

Computer modeling can be done to account for the dynamic or time-varying nature of the follow-up. Longitudinal data, of course, are needed not only of the outcome but also of conditional or mediating variables, along with predisposing characteristics that are both static (e.g., race, personality) and dynamic (e.g., employment, drinking). One popular analytic tool for this purpose is Generalized Estimating Equations (GEE; Laing & Zeger, 1986). It derives an equation for the pooled follow-up intervals, and the correlation between those intervals, using a series of equations predicting the outcome of each separate follow-up interval outcome (thus producing consistent standard errors). GEE also accommodates so-called censored data of individuals who do not complete the full follow-up.

The more complex issue is how to account for the broader context, while controlling for the differences in program dropouts and completers. One approach, also developed in the public health field, is Instrumental Variable (IV) analysis (e.g., Bollen, Guilkey, & Mroz, 1995). In this structural equation approach, predisposing characteristics are entered in an equation to predict dropout along with a set of contextual “instrumental variables” that uniquely distinguish dropout (e.g., victim services, referral source, and perceived sanctions). This “dropout” equation is then entered into a second equation predicting reassault, along with predisposing variables and IVs unique to reassault (e.g., additional services, unemployment, and police domestic violence arrests). This approach also accounts for potential

“unobservable” characteristics of the batterers (e.g., stress, motivation, and attitudes). It does, however, have assumptions that are difficult to meet and addresses only cumulative outcomes (not dynamic ones).

A more extreme position views program outcomes as the result of a “social construction” in which researchers, program staff, and subjects all participate (e.g., [Guba & Lincoln, 1989](#)). Researchers asking questions, collecting information, and interpreting information participate in an interactive social process that creates their own “definition of the situation.” Evaluation is, consequently, not an objective or purely scientific process that produces unbiased and conclusive results. In this view, a program evaluation is a process with a subjective outcome. The ideal evaluation, in this view, would be a kind of case study that describes the entire process, changes, and perceptions, and acknowledges the subjective nature of all of this. These approaches sometimes fall under the rubric of “postmodern” thought. Other times they may be referred to by their method: qualitative, ethnographic, context-specific, action-oriented, realistic, social constructionist, or fourth generation evaluation (e.g., [Guba & Lincoln, 1989](#)). These alternative approaches have gained a strong foothold in graduate schools of education and social work and resonate with many practitioners.

The main problem is that the subjectivity of these evaluation approaches makes it hard to generalize results and apply them to policy. Some of this problem may be partially addressed through a kind of compromise in which so-called scientific research is conducted in collaboration with practitioners. “Collaboration” has become a popular aspect of evaluation, but it can be difficult to implement and utilize in a way amenable to both research and practitioners, as recent case studies illustrate ([Edleson & Bible, 2000](#); [Gondolf, Yllo, & Campbell, 1997](#); [Murphy & Dienemann, 1999](#); [National Violence Against Women Prevention Research Center, 2001](#)). These same collaboration studies suggest some positive directions, however. Although the studies offer no unified definition of collaboration, they agree that collaboration is helpful in identifying program context and interpreting results.

All of these issues suggest the following ideal for batterer program evaluations. Evaluations need to address a dose-response approach through modeling techniques that simulate an equivalent control group, consider dynamic and conditional outcomes, and recognize contextual factors and collaborative interpretations. The analytical procedures, of course, need to be sufficiently sophisticated to account for such a complex model. As a reanalysis of the Spouse Assault Replication Program (SARP) shows, a different analytical procedure may produce different results ([Maxwell, Garner, & Fagan, 2001](#)).¹ Evaluations of this sort are admittedly costly and sometimes impractical to achieve. It is difficult enough to conduct some of the more basic evaluation designs in a criminal justice setting. The ideal offers, in any case, a set of considerations to weigh in assessing the results of evaluations that are available and in formulating new ones for the future.

¹ SARP was the multiple replications of the Minneapolis Police Study, an experimental evaluation assessing police arrest. The separate replications had some contradictory results, but the reanalysis of the combined data reaffirmed the effectiveness of police arrest in reducing reassaults.

2.4. Previous evaluations and current meta-analysis

2.4.1. Meta-analyses and effects

Over 40 batterer program evaluations have been published in academic journals, and many more are available as agency reports. Reviews of the batterer program evaluations show 50–80% of the program completers to be nonviolent at the end of a 6-month to 1-year period, according to their female partners (e.g., Edleson, 1996; Gondolf, 1997a; Rosenfeld, 1992; Tolman & Bennett, 1990). The reduction of other forms of abuse is less clear, but one study showed that only about 40–50% of the participants had stopped their terroristic threats at a 6-month follow-up (Edleson & Syers, 1990). It may be that some men displace their physical abuse to heightened verbal and psychological abuse. The batterer program “success rates,” while compromised by high dropouts and incomplete follow-ups, are comparable to those in drunk driving, drug and alcohol, sex offender, and check forging programs (see Furby, Blackshaw, & Weinrott, 1989; Hubbard et al., 1989; Kassebaum & Ward, 1991; Schare, 1992). In a social science “court,” most all these batterer program evaluations would, however, be dismissed on technicalities or as circumstantial evidence (see Quinsey, Harris, Rice, & Lalumiere, 1993). They are compromised by selection bias, low response rates, short follow-up periods, no or weak control groups, and no calculations of effect size.

The effectiveness of batterer programs has been addressed in a more comprehensive way using meta-analyses of both experimental and quasi-experimental evaluations (Durlak & Lipsey, 1991).² These analyses statistically summarize the magnitude of the effect attributable to the group of programs considered in the analyses. Cohen (1977) developed statistical calculations for effect size as an alternative to the dependence on and misuse of *P* values. In the case of batterer programs, the coefficient for “effect size”—most commonly Cohen’s *h*—considers the difference in reassault rates between the program and the control or quasi-control group adjusted for the sample size. Three meta-analyses have assessed the effect size of batterer program evaluations that include some sort of control group (Babcock & Robie, in press; Davis & Taylor, 1999; Levesque & Gelles, 1998). Unfortunately, the research methods, program implementation, and program context influence the results of these analyses and

² Meta-analysis has become a popular means of summarizing a body of empirical research. It appears more scientific than descriptive reviews, and claims to eliminate the bias that researchers may bring to their individual studies or literature reviews. It also serves as a way to referee, with an apparent objectivity, disputes and conflicts in a field, and helps to temper inflated claims of program “success.” While meta-analysis definitely has some contributions to make, especially in comparing program effectiveness across evaluations, it can be misleading for a number of reasons. The conceptual problems with design and implementation are often difficult to access without an intricate knowledge of the field, and consequently they may be neglected by the meta-analysis or left to caveats in the conclusion. Meta-analysis results must still be interpreted with the subjectivity of perspective and context. Meta-analysis tends to decontextualize research studies by detaching them from the period, circumstances, and issues in the field at the time of each evaluation. It also leaves out the discourse among researchers and practitioners about the meaning and utility of the research. Not only is the complementary accumulation left out of the meta-analysis but so are many of the studies that do the complementing.

pose several caveats (Wilson & Lipsey, 2001). The effect size is not, therefore, a “bottom line” of program success or failure independent of interpretation, although it is sometimes used that way in these analyses.

The most current meta-analysis selected 15 of the most methodologically sound batterer program evaluations and included the three recent experimental evaluations discussed below. It found, similar to the other previous meta-analyses, only a small effect size for batterer programs overall (Babcock & Robie, *in press*). Effect sizes (computed with the *d* statistic) were .10 for the experimental designs ($N=8$ effects), .41 for the quasi-experimental evaluations, and .24 for all of the studies combined ($N=20$ effects; $CI=.15-.33$) based on women’s reports. The authors raise several caveats, however, and acknowledge the variability in the quality of the evaluations, confounding effects of the criminal justice system, and moderate effects of several programs lost in the averaging. Two programs, moreover, produced effect sizes of .70 and three others in the .40–.50 range. The authors conclude with the cautionary note: “One of the greatest concerns when conducting a meta-analysis is the ease at which the ‘bottom line’ is recalled and the extensive caveats for caution are forgotten or ignored.”

2.4.2. Meta-analyses caveats

Statisticians have raised additional concerns with the bottom line use of meta-analyses (e.g., Cohen, 1994; Jones & Scharfstein, *submitted for publication*; McCartney & Rosenthal, 2000). Despite the popularity of reporting effect size as small, moderate, or large, Cohen (1994) and others argue this practice is frequently applied in a misleading way. Small effects are still significant effects and may be a sufficient endorsement for a program, since the guidelines used for effect sizes are generally too stringent (Mann, 1990). The nature of the expectations for the program, characteristics of the program participants, available alternatives to the program, and costs of the program may make some programs with a so-called small effect worthwhile. For example, we would expect a smaller program effect with men repeatedly arrested for crimes than a program with clients who voluntarily sought out treatment. Moreover, most meta-analyses show some programs with at least a moderate effect and others with slight effects, as did the most recent meta-analysis of batterer program evaluations.

There is one other caveat raised with meta-analyses in general. The meta-analyses are almost entirely based on single-site evaluations. In fact, one of the main functions of meta-analysis is to pool the effects of several single-site evaluations. As a result of variation in measurement, design, and context, program outcomes and effects may be distorted. This becomes most apparent in controlled multisite studies or pooled replication studies that often produce a different result than the accumulation of previous single-site studies. The most prominent example of this is Project Match, the multisite study of alcohol treatment funded by the National Institute of Alcohol Abuse and Alcoholism (NIAAA; Project MATCH, 1997). The accumulation of single-site studies suggested treatment effects according to patient type, but the multisite study established counter evidence. The multisite depression treatment clinical trials, funded by the National Institute of Mental Health (NIMH), produced some similar reversals (Elkin et al., 1989).

2.4.3. *Experimental limitations*

The methodological limitations of the experimental designs may be a substantial reason for the slight effect size found in the meta-analysis. As a special issue of *Crime and Delinquency* argues, the experimental evaluation of batterer programs should get the most weight, since their designs are the most rigorously scientific (Boruch et al., 2000; Dunford, 2000a). A careful look at the three recent experimental evaluations exposes, however, implementation problems that compromise their results (for details, see Gondolf, 2001). Both the New York City (Davis, Taylor, & Maxwell, 1998) and South Florida (Feder & Dugan, 2002) evaluations had substantial problems implementing the random assignment. As much as 30% of the potential cases were overridden at one site, and the point of random assignment was moved twice at the other. Moreover, legal opposition to random assignment disrupted the South Florida evaluation (Feder et al., 2000). The batterer program in New York was changed during the subject recruitment.

There were also major problems in implementing the follow-up. The response rate of female partners was only 20% in the South Florida evaluation, which led to the use of confounded probation records as the outcome source. The New York City evaluation used private investigators to track down some nonrespondents, which may have affected their disclosure. The low program completion—as low as 40% for the South Florida program—exacerbated the intention-to-treat distortions of the experiments, as well, and raises questions about the program operation in general. (Completion rates at programs in our multisite evaluation were 55–70% varying with the required length of the program (Gondolf, 1997b).)

A “clinical trial” experiment at a Navy base showed no effects for batterer counseling compared with other options such as safety planning and intensive monitoring (Dunford, 2000b). This evaluation is difficult to generalize to the civilian public, however, because of the different characteristics of the Navy men, and the heavy and certain sanctions for reassault while under Navy supervision. Moreover, the comparative options in the Navy experiment, such as the monitoring and safety planning, are incorporated as part of the intervention system in most civilian batterer programs rather than being alternatives. Consequently, the results of the previous experiments, although instructive, must be viewed and applied with caution.

In summary, it is difficult to make a categorical recommendation about batterer program effect given the methodological problems of many of the evaluations and the limitations of effect-size statistics. The meta-analyses do, however, raise suspicions about excessive claims of success; we have no substantial evidence that most programs are effective or that any programs are highly effective.

3. A multisite evaluation of batterer intervention

3.1. *Multisite design*

We developed a multisite evaluation in an effort to address some of the conceptual issues and methodological shortcomings of previous evaluations (Gondolf, 1997b, 1999a, 2000d,

2001). This evaluation had the advantage of substantial funding from the Center for Disease Control and benefit of both clinical and research advisors. In our opinion, the results of the evaluation indicate that at least some programs are effective in stopping assault and abuse and that batterer intervention in general shows some promise. The evaluation also suggests that the predominant gender-based cognitive–behavioral approach to counseling may be appropriate for the majority of men. The slight effects shown in other evaluations may be the result of poorly implemented programs or intervention systems.

Our multisite evaluation employed a naturalistic comparative design across sites, with quasi-experimental studies within the sites. We identified programs in four cities that represented a range in program format and system components. The four interventions were based on “well-established” batterer programs that used the prevailing gender-based, cognitive–behavioral approach. The intervention systems ranged, however, from a streamlined system of (1) pretrial referral to batterer counseling at a preliminary hearing, (2) additional referrals for court-identified alcohol or psychological problems, (3) 3 months of instructional group counseling, and (4) periodic court review of program compliance. At the other extreme was a comprehensive system based on (1) postconviction referral, (2) individual assessment, (3) in-house alcohol and psychological treatment, (4) 9 months of discussion-oriented group counseling, and (5) monitoring of compliance by probation officers. The remaining two systems included a 3-month discussion-oriented program and a 5 1/2-month instructional program, both with postconviction referrals.

The evaluation consisted of a 4-year follow-up, starting at program intake, with periodic interviews of 840 men and their female partners. During 1995, research assistants recruited program participants from the four sites ($N=840$) and administered a uniform set of background questionnaire, personality inventory (MCMI-III; Millon, 1994), and alcohol test (MAST; Selzer, 1971). The sample size was reduced in the extended follow-up (15–48 months) to 618. The participants who were voluntary and recruited in the first 2 months were deleted. This deletion furthered the focus on the court-referred men and made the follow-up less costly. (The reassault rates presented here are those computed for the court-referred men alone.) A variety of qualitative and quantitative outcomes, and conditional or time-varying variables, was assessed at 3-month intervals for 4 years. The response rate with the female partners of the program participants was approximately 70% for the first 30 months and 60% for the full 4 years (for human subjects and tracking procedures, see Gondolf, 2000b).³

The principal outcome of reassault was based on reports from the men’s initial and new female partners, and confirmed with an analysis of police reports and men’s self-reports (Heckert & Gondolf, 2000a,b), a coded qualitative analysis of women’s narratives (Heckert,

³ The most persistent concerns about outcomes in batterer program evaluations are the reliability of women’s reports. Don’t the women deny or withhold the truth of being reassaulted? We believe that disclosure was facilitated by our frequent and long-term contact, the funnel questioning that allowed women to tell their “story,” and debriefing interviews that inquired about obstacles or barriers. All of the women had, moreover, already officially reported (or had someone report for them) their initial assaults and had to appear in court to testify about that assault. Therefore, reporting anonymously to a researcher after the court appearance was not an unfamiliar or foreboding task for most. In fact, many of the women were very eager for someone to know about their partners’ subsequent behavior and to offer information that might improve the interventions for other women.

Matula, & Gondolf, 2000), and a Capture–Recapture analysis of self-reports and arrest records (Gondolf, Chang, & LaPorte, 1999). Several attrition analyses, including one using Heckman (1979) regressions, revealed no significant response bias (Jones, 1998). The scope of the data, size of the pooled sample, and longitudinal follow-up enabled us to conduct some more complex analyses of the dynamic outcome (Jones & Gondolf, 2001) and the contextualized program effect (Gondolf & Jones, 2001). We also monitored the counseling sessions and conducted periodic observations and interviews about the intervention system and community context (see Gondolf, 2002a, chap. 4).

3.2. Reassault rates

We first considered the reassault rates for both program completers and dropouts at the four sites. These rates offer some assessment of the policy of court referral to these batterer programs, regardless of program dose or compliance. The cumulative reassault rate (i.e., from intake through the follow-up period) was 32% for the first 15 months, 38% for 30 months, and 42% for the full 48 months, according to the women’s reports. The 4-year rate increased to 47% when the women’s reports were adjusted for possible underreport using arrest records and men’s reports. The Capture–Recapture analysis estimated a 49% reassault rate for the full sample at 4 years.

A more positive picture emerges when reassault is considered *retrospectively* from the end of the follow-up period, as is increasingly being done in alcohol treatment evaluations. These rates allow for the intervention to “take effect,” so to speak, before counting the reassaults. At the 30-month follow-up, less than 20% of the men had reassaulted their partner in the previous year; at the 48-month follow-up, approximately 10% had reassaulted in the previous year. Moreover, over two-thirds of the women said their quality of life had improved and 85% felt very safe at both these follow-up points.

Our longitudinal data enabled us also to examine the trend of reassault over time. There is a marked deescalation of reassault (Gondolf, 2000d), and also of other forms of abuse (Gondolf, Heckert, & Kimmel, 2002), rather than an escalation after counseling as a result of a short-term deterrence effect. Nearly three-quarters of the reassaults during the 15-month follow-up period occurred within the first 6 months—that is, when the men were supposedly in the program. This finding suggests a need to better supervise and contain men immediately after program intake, rather than focus primarily on “after the program.”

Considering the severity and duration of the men’s previous abuse and the compounding alcohol and psychological problems of many of the men (see Gondolf, 1999b), this level of reassault might be considered an accomplishment. According to a cost analysis by a health economist (Jones, 2000), the cost for one man per counseling session was only US\$20–30, and half of that cost was covered by the program participants. The reassault outcomes and low session cost suggest that court referral to batterer programs may be acceptable, even if it is not substantially more effective than the other alternatives of short-term jailing or open probation.

We cannot be sure, however, if court referral is more effective than other court actions, since we do not have a direct comparison to these other actions. A previous uncontrolled study comparing outcomes for jailing, program dropout, and program completion showed

lower rearrests rates for program completers (Babcock & Steiner, 1999), as does our comparison of arrest rates for different court actions at one of our sites (Gondolf, 1998). At the minimum, the court referral offers an intensified supervision through the program and a weekly reprieve for the men's partners.

3.3. Program effect

The second and more debated question is whether the counseling is effective over no counseling—that is, is there a program or treatment effect beyond the arrest and court appearance? When we compared those men receiving a minimum dose of counseling (less than 2 months) with those who completed 2 months or more, we found a 50% reduction in reassault for the “completers” (36% of the completers reassaulted vs. 55% of the dropouts). This reduction amounts to an effect size of .48 (Babcock & Robie, *in press*). The site with the streamlined system was deleted from this analysis because the quasi-control group was contaminated by the court-review process. Most dropouts at this particular site received compensating interventions that contained their potential reassault (for documentation, see Gondolf & Jones, 2001). The reduction in reassault is even higher when controlling for the men living with their partners (66% reduction: 40% vs. 67%). Many men drop out because of no longer being with their partner and not seeing the relevance of counseling.

Different characteristics between the dropouts and completers may, of course, account for the apparent reduction attributed to the program. When controlling for demographic, relationship, personality, and behavioral characteristics, in a logistic regression (or naïve analysis), program completion remains significant with a reduction of 18% in reassault.⁴ To account for the endogenous nature of dropout, we also conducted the form of structural equation analysis called IV analysis. According to this approach, program completion produced an effect size of .44–.64, depending on the model specification (Gondolf & Jones, 2001). This effect was confirmed in a “Propensity Score analysis” showing an effect of .50 for court-referred men (Jones et al., *in press*). As mentioned previously, both these analytical approaches are well-established means for assessing effects in the public health field where experimental designs are most often impractical or inappropriate. The moderate program effect is, moreover, confirmed by circumstantial evidence from a deterrence analysis showing little impact of perceived sanctions (Heckert & Gondolf, 2000c), attribution of change to the programs, use of avoidance techniques from the programs (Gondolf, 2000a), and program recommendations from both men and women (Gondolf & White, 2000).

3.4. System effect

We expected the reassault rates to be the lowest among intervention systems with longer and more comprehensive programs for a number of reasons. One, our book based on

⁴ This analysis is more conservative than the previous uncontrolled effect size because it uses 3 months as the cut-off for dropouts rather than 2 months. Three months were used to ensure a larger distribution of dropouts for the multivariate analysis.

interviews and observations of batterer programs across the country promoted a comprehensive approach (Gondolf, 1985). According to feminist and social learning theories, men are, furthermore, socialized by a male-dominated society to exert power over women and reinforced by a violent culture (e.g., Taubman, 1986). It would take a long and concentrated effort to “unlearn” this socialization. However, the reassault and related outcomes across the four sites were strikingly similar (Gondolf, 1999a). When we controlled for demographics, relationship status, personality, and previous behavior, we still found no significant site effect on reassault.

The lack of a site effect was confirmed in the IV analysis which tested for a program effect (Gondolf & Jones, 2001), and in the GEE analysis identifying predictors of reassault (Jones & Gondolf, 2001). At the 15-month follow-up, there was a slight decrease in *severe* reassault in the longer programs, but this was not corroborated by other indicators and disappeared at the cumulative 30-month follow-up (Gondolf, 1999a; Gondolf, 2000d).⁵ A similar result regarding the effectiveness of program length has appeared in evaluations of sex offender treatments. Longer treatment programs have not been shown to be more effective than shorter treatment programs (Marshall & Serran, 2000).

There are several possible explanations for the lack of a “site effect.” One explanation reflects the assumptions of brief therapy and managed care: The men who are going to change begin to do so within 3 months of counseling, and those who might benefit from longer counseling tend to drop out. Indeed, 3 months represented not only the duration of two of the programs, but also a threshold of dropout for the longer programs (Gondolf, 1997b). Another explanation is the possible uniqueness of each intervention system. The NIMH multisite depression study found an interaction between site and counseling approach that appeared to be related to cultural and resource differences among the sites (Elkin et al., 1989). The batterer intervention systems in our evaluation similarly evolved in response to a distinctive set of circumstances, resources, expectations, and personalities. For example, the site of the longest and process-oriented program had more therapists per capita than all but one other city in the country. This circumstance may account for the batterer program participants at this site being the most likely to have previously received counseling and expect longer, process-oriented batterer counseling.

The most compelling explanation for no site effect, however, may be the compensating influence of system components. The streamlined system operated much like the increasingly popular drug courts. Its court supervision of the cases ensured a swift and certain response to noncompliance (Gondolf, 2000c). Under the pretrial referral, the men entered the program in an average of 2 1/2 weeks after arrest, as opposed to several months at the postconviction systems, and they had to reappear in court periodically to confirm their program attendance. This system dramatically reduced no-shows (from 30% to 5%) and sustained a high completion rate of 70% despite the coerced attendance. The system appears to matter.

⁵ The distribution of psychological problems and previous behavior was strikingly similar across the four sites, but socioeconomic status was lowest for the streamline site and greatest for the comprehensive site (see Gondolf, 1999b).

3.5. Repeat reassaulters

A secondary finding points to areas for improvement. Approximately a quarter of the men (about half those who reassaulted during the four-year follow-up) reassaulted their partners more than once (Gondolf, 2000d). Most of these men began their reassaults shortly after program intake, and were responsible for over 80% of the injuries. They were, not surprisingly, more likely to have been severely violent in the past and to have been arrested or treated previously. However, their personality profiles did not distinguish them from men who reassaulted only one time, and men who did not reassault (Gondolf & White, 2001). Their patterns of violence, based on coding of the women's narratives, also did not distinguish these particularly dangerous men (Gondolf & Beeman, 2003). The analysis of women's narratives did, however, reveal distinct patterns and trajectories of violence that warrant further consideration. Calibrated identification of violence patterns might improve the prediction of further violence. This possibility makes sense given that the gross indicators of antisocial behavior continue to be correlated with program outcome—"past violence is the best predictor of future violence" (e.g., Dutton, Bodnarchuk, Kropp, Hart, & Ogloff, 1997).

In summary, the repeat reassaulters did not appear as a distinct "batterer type." The most distinguishing factor was the lack of response to these men. Their partners were less likely to take action, possibly out of fear or subjection; and further arrests, protection orders, and treatment were not as likely, as a laboratory study found with more antisocial men (Jacobson & Gottman, 1998). More extensive case management and systematic victim contact might help expose repeated reassaults, and certain and decisive intervention for a initial reassault might make a difference on the outcome. It would likely reduce repeated reassaults.

Another option is to attempt to identify these "high-risk" men at program intake, as several risk assessment instruments attempt to do (for overviews of these instruments, see Dutton & Kropp, 2000; Roehl & Guertin, 2000). Risk assessment instruments have been shown to improve prediction over clinical judgment, but their predictive power still falls far short of what could be considered reliable. We attempted to identify predictors for reassault and especially for repeat reassault. Static intake variables showed little predictive power for reassault during the crucial first 15 months (Jones & Gondolf, 2001). The only significant intake predictors were severe psychological problems, previous severe abuse, as other research of violent men has found (Walters, 2000). The only substantial predictor was drunkenness during the follow-up. (Those men identified by their partners as "drunk," at least once during a follow-up interval, were four times more likely to reassault.) We also found at least a weak association between accompanying alcohol treatment and a reduction in reassault; however, the treatment varied considerably in extent and intensity and was not sustained with many of the most violent men.

This latter finding lends support to the recurrent association of alcohol misuse to woman assault. The relationship between alcohol and assault is, however, a complex one, as numerous reviews of the topic have indicated (e.g., Gondolf, 1995). We were able to demonstrate the separation between "drunkenness" and reassault by lagging the drunk-

eness in the analysis and also with a separate variable for “drinking prior to the assaultive incident.” The drunkenness, therefore, appears to be more of a matter of lifestyle, rather than a direct cause. Interestingly, other research has shown that the association of drinking to reassault disappears when male attitudes of dominance are controlled (Johnson, 2000). The heavy drinking and reassault may both be symptoms of the same underlying attitudes or lifestyle. Some researchers have, more broadly, implicated unstable lifestyles as predictive of reassault by violent offenders in general (Hanson & Wallace-Capretta, 2000).

3.6. *Additional prediction studies*

We conducted further predictive analysis using multiple outcomes (i.e., no abuse, nonphysical abuse, threats, one reassault, and repeat reassault) in an effort to improve predictive power and identify additional predictors (Heckert & Gondolf, 2001). Again, previous reassault and a few demographic variables distinguished severe reassault, but correct classification, although improved, is still weak. Our multinomial analysis using the five possible outcomes correctly classified only 46% of the total cases. However, it did correctly classify 59% of the repeat reassault cases—the men of greatest concern. While this is a substantial improvement over the correct classification of reassault in the dichotomous outcome of reassault versus no reassault (41%), the classification rate for repeat reassaulters is still relatively weak for two reasons. One, 40% of the actual repeat reassaulters were “false negatives”—that is, they were classified as not being repeat reassaulters when they actually were. Two, 50% of those classified as being repeat assaulters were “false positives”—that is, they were classified as repeat reassaulters but were *not*. The correct classification rates for the other outcomes were: 39% for “no abuse,” 63% for “controlling or verbal only,” 35% for “threats,” and 10% for “one-time reassault.” The strongest predictors were consistent with the prevailing ones in predictive studies of violent offenders: prior severe abuse, prior criminality, antisocial personality, and substance abuse (see Gendreau, Little, & Goggin, 1996).

Items used in various risk assessment instruments did not improve the prediction. We were able to simulate three of the leading batterer risk assessment instruments [i.e., the Kingston Screening Instrument for Domestic Violence Offenders (K-SID), the Spousal Assault Risk Assessment (SARA), and Danger Assessment Scale (DAS)] with our comprehensive intake data. The simulated instruments did not perform any better than the previous analysis with individual intake risk markers (Heckert & Gondolf, 2001).

Our initial expectation was that batterer types or stages would be associated with outcomes, as a consolidation of batterer-type research suggests (Holtzworth-Munroe & Stuart, 1994). Our own previous research identified batterer behavioral types based on reports from a sample of battered women in Texas shelters (Gondolf, 1988), and a theoretical article identified “stages of change” based on Kohlberg’s developmental theory and their implications for program curriculum (Gondolf, 1987). The current batterer types, however, have not been established as discrete diagnostic or predictive categories. They ultimately need to be tested in longitudinal clinical samples to assess their relevance for

intervention. This has been done in NIAAA's elaborate test of specialized treatment for different types of alcoholics (Project Match, 1997). This multisite study concluded that "patient-treatment matching . . . adds little to enhance outcome treatment" (Gordis, 1997, p. 6).

To explore the utility of batterer types, we created personality-based types in two ways: (1) with a cluster analysis of MCMI-III subscales, and (2) with a classification of the MCMI-III profiles based on interpretative guides (White & Gondolf, 2000). Both approaches produced "types" that reflected tendencies suggested in the prevailing typology (Holtzworth-Munroe & Stuart, 1994). The most remarkable part was the percentage of men who showed no "psychopathology" at all, and the diversity of profiles overall, as found with clustering using the MCMI-II (Hamberger, Lohr, Bonge, & Tolin, 1996). We did not find any evidence to substantiate a prevailing "abusive personality" with underlying tendencies of a borderline personality (Dutton, 1998).⁶

Our types are not, however, as complex as the personality/behavioral types in the most current typology research (Holtzworth-Munroe, Meehan, Herron, Rehman, & Stuart, 2001). The violent and criminal behavior incorporated in the more complex batterer types may make the types appear predictive, as has been suggested with measures of psychopathy (Skeem & Mulvey, 2001). Antisocial behavioral indicators (e.g., prior violence and criminality) and antisocial personality subscales in general correlate the most strongly with outcome (e.g., Dutton et al., 1997; Heckert & Gondolf, 2001). (For a discussion of the limitation of stages of change as the basis for another batterer typology, see Sutton, 2001.)

The strongest and most consistent predictor in our multisite evaluation was the women's perceptions (i.e., likelihood of reassault, how safe you feel), echoing the results of a similar prediction study (Weisz, Tolman, & Saunders, 2000). These findings lend support to an alternative approach to prediction, ongoing risk management. Violence research in the psychiatric field (Monahan et al., 2001) and recent Secret Services studies (Fein, Vossekuil, & Holden, 1995) promote this approach over the use of static risk assessment instruments at program intake. Ongoing risk management assesses short-term risk periodically, introduces interventions and treatments in response to any immediate risk, reassesses the risk again, and repeats the process. In this way, it attempts to account for conditional or time-varying factors and the dynamic interplay of a variety of cues, like those that female partners observe and respond to (Langford, 1996).

⁶ The conception of an abusive personality obviously warrants further conceptual and empirical testing. Our failure to replicate it with the MCMI is, no doubt, related in part to instrumentation and contrasting conceptions of borderline tendencies. The MCMI developers, for instance, argue that the majority of psychological disorders have underlying borderline tendencies, and that borderline tendencies may by themselves be assessing psychological dysfunction in general (Craig, 1995; Millon, 1994). Also, the difference in the prevalence of borderline tendencies could be the result of different samples of batterers. Our sample was comprised mostly of lower income and racially diverse men, as opposed to the Dutton samples based on a more homogenous and employed Canadian population. Antisocial and narcissistic tendencies were predominant in our sample that included nearly half of the men with prior arrests.

4. Program and evaluation implications

4.1. Program implications

The most surprising finding in our multisite evaluation is that the vast majority of men referred to batterer counseling appear to stop their assaultive behavior and reduce their abuse in general. The batterer programs, in our evaluation, appear to contribute to this outcome—there is a “program effect.” Referral to the gender-based, cognitive–behavioral programs, moreover, seems to be appropriate for the majority of men. This finding echoes the evaluation outcomes and program recommendations for the treatment of violent offenders in general, including sex offenders (e.g., Marshall & Serran, 2000; Rice, 1997). Comparison studies of counseling approaches are obviously still needed to confirm this current interpretation and explore the role of the components within the cognitive–behavioral approach. To date, there has been only one clinical trial of different counseling approaches for batterers.

A small study in Wisconsin compared men randomly assigned to gender-based, cognitive–behavioral counseling and to psychodynamic counseling (Saunders, 1996). The outcomes of the two approaches were very similar. There was, however, an interactive effect between personality tendencies and counseling approach. The men with more antisocial tendencies did slightly better in the cognitive–behavioral counseling, and those with depressive tendencies appeared to do better in the psychodynamic counseling. The cognitive–behavioral approach continues to hold some precedence, in part, because of the antisocial tendencies that tend to typify court referrals, as was the case in our sample. It also tends to be more efficient and is less costly to implement.

Our analysis of personality profiles supports this “one size fits most” implication (White & Gondolf, 2000). A substantial portion (56%) of the men showed no evidence of personality disorders or major psychological problems, and most of those who did were suitable candidates for cognitive–behavioral counseling, according to the profile interpretative guides and treatment convention (e.g., Craig, 1995). Moreover, the most consistent elevation on the MCMI was narcissism (Gondolf, 1999c), as research on violent offenders in general suggests (Baumeister, Smart, & Boden, 1996). One study of counseling with imprisoned violent offenders shows that other counseling approaches may inadvertently reinforce the men’s narcissism and actually increase their violence (Rice, 1997). Another study suggests that anger, or more emotive aspects addressed in psychodynamic approaches, do not distinguish more severe violent offenders from less severe violent offenders and may not be the most appropriate target for counseling (Loza & Loza-Fanous, 1999a).

Violent offenders obviously have differences in personality and behavioral patterns, but they may have overarching commonalities that are addressed in the prevailing counseling approach. Or, as proponents of brief and solution-oriented therapy argue (Budman & Gurman, 1988), the means for stopping certain behaviors may be independent from its “causes.” That is, why a man abuses others may not be the same as how he stops the abuse. There is, nevertheless, a portion of men who should be screened for severe psychiatric disorders and referred to complementary or alternative treatment for those disorders.

4.2. System coordination

Our findings shift the focus from curriculum change or treatment diversification toward program structure and system coordination. In terms of program structure, program intensity rather than extent warrants more attention. More might be done to monitor and contain men during the first few months after program intake, when men are most likely to first reassault. Much like the move toward intensive outpatient treatment in the alcohol field, some men might be required to attend three to four long sessions per week, rather than the usual once-a-week sessions of 1.5–2 h. The men who have reassaulted, or have severely assaulted their partners in the past, are logical candidates for these sessions. Programs might also offer more continuous support to the men's female partners through women's advocates or caseworkers. The outreach to the women could be substantially improved within the intervention systems of our study. While about a third of the batterers' female partners reported some contact with battered women's services within the first 3 months of program intake (beyond contact with a legal advocate), only 8% of the women had any contact in the next 12 months (Gondolf, 2002b). Most of this later contact was "reactive" in that it was a response to additional assaults. Consequently, women services were not associated with a reduction in reassault (Jones & Gondolf, 2001). Additional support might develop ongoing risk management that would bring additional intervention and assistance as appropriate, especially in cases of repeat reassault. It might also help women more assertively access and benefit from available resources and services.

Several areas need attention within the intervention system as a whole. More men need to be moved swiftly into batterer programs. The effectiveness of programs is apparently undermined by long delays between arrest and court disposition and by the large number of program no-shows and dropouts. The large number of men lost to dismissed and withdrawn cases contributes to this problem. Also, the court response to noncompliance after program intake needs to be more swift and certain. Pretrial referral and "fast-track" prosecution, in specialized domestic violence courts, are two ways to address these "before" and "after" program problems. Pretrial referral appeared particularly effective at one of the sites in our study, although it meant no criminal record under pretrial practices (Gondolf, 2000c). This approach operates similar to the increasingly popular "drug courts" used to speed the response and supervision of drug-related cases (Belenko, 1998; Gottfredson, Najaka, & Kearley, 2003).

4.3. Additional concerns

One of the lingering questions about batterer programs, assuming some programs are effective, is how to get more nonarrested men into them. Most programs not only receive a very small portion of the men arrested for domestic violence, but also arrested men comprise a very small portion of men who abuse women. Obviously, more needs to be done at the community level to encourage men to seek help for their abuse of others, much as is done with men who misuse alcohol. There is, however, a reduced stigma for seeking alcohol treatment and even a positive status for being "in recovery." Alcohol misuse is more clearly

detrimental to the drinker as well as to others. Some different counseling approaches or formats may also be needed for nonarrested men. A subsample of “voluntary” participants in our evaluation tended to drop out of conventional batterer treatment and reassault their partners (Gondolf, 1997b). They were also more likely to have severe psychological problems, particularly depression.

An additional consideration is the possible need for specialized counseling for African American men and Latino men (see Carrillo & Tello, 1998; Williams, 1994). Our review of the research shows cultural issues that may account for higher dropout and poorer outcomes among these men (Gondolf & Williams, 2001). However, there is little research that documents poorer outcomes in conventional programs or improved outcomes in culturally specific programs. One preliminary study of specialized batterer counseling shows African American men having more favorable impressions of culturally specific counseling ($N=41$; Williams, 1995). However, the African American men in our multisite evaluation were no more likely to reassault than their white counterparts. To examine this issue more systematically, we have begun a clinical trial of specialized counseling for African American men arrested for domestic violence. This evaluation, funded by the National Institute of Justice, compares conventional batterer counseling that is racially mixed, conventional batterer counseling with only African Americans, and culturally specific counseling with only African Americans. Preliminary results show that the culturally specific counseling improves at least program completion (Gondolf, *in press*).

As a recent review of domestic violence and sex offenders suggests (Saunders & Hamill, *in press*), there may be some merit in exploring both the programmatic and evaluation developments in related fields. The sex offender field has a proliferation of risk assessment instruments and several meta-analyses showing some effectiveness from cognitive-behavioral treatments (Marshall & Pithers, 1994; Marshall & Serran, 2000). It has also developed more holistic treatments and community surveillance of offenders. The alcohol treatment field has, moreover, advanced treatments for dual diagnoses while continuing confrontational sessions and intensive outpatient treatment. It puts some emphasis on spiritual development and follow-up mutual support groups (i.e., Alcoholics Anonymous) that have yet to take hold in the domestic violence field.

4.4. *Evaluation implications*

Previous evaluations and our own multisite evaluation raise a few cautions for the wave of evaluation research. It may be helpful to educate program staff, policy makers, and funders about the limitations of most evaluations. While evaluations provide useful feedback and guideposts, they require substantial cost and sophistication to establish definitive results. In fact, categorical recommendations (e.g., programs “work” or “do not work”) are seldom achieved. We may need to think about pooling resources to conduct more multisite evaluations with clinical trials of program structure and system components in different contexts rather than proliferate a host of limited single-site evaluations. Meta-analysis can help appraise a variety of single-site evaluations, but this approach does tend to decontextualize the evaluations and still requires interpretation.

At a minimum, evaluation might consider the influence of system components on outcomes. We do not know, for instance, the effectiveness of a comprehensive, 9-month program, coupled with the swift and certain response of a “drug-court type” referral and supervision. We do not know the trade-offs between pretrial referral versus fast-track prosecution, or between court review versus intensified probation. We do not know if these options are unique to certain court jurisdictions and community settings. The National Institute of Justice has launched a multisite demonstration project of more highly coordinated, and even integrated, intervention systems. The evaluation component attached to this project may offer some answers in this regard.

Finally, there remains the larger question of how best to develop counseling programs and intervention systems. The administration of programs varies tremendously from a clinician-led set of groups to agency-based counseling sessions. Some communities have a scattering of programs competing against each other, and some have one or two programs that maintain a monopoly over the counseling. Some programs have established court liaisons and case managers to help coordinate the intervention components; others rely on specialized probation departments to take this responsibility. It is clear, from our observations, that different administration has different effects on the system and ultimately on the program outcomes (see Gondolf, 2002a). It is something that also warrants more attention and development. To make some of these most basic developments requires a means to do so.

References

- Adams, D. (1989). Feminist-based interventions for battering men. In L. Caesar, & K. Hamberger (Eds.), *Treating men who batter: Theory, practice and programs* (pp. 3–23). New York: Springer.
- Aldarondo, E. (2002). Evaluating the efficacy of interventions with men who batter. In E. Aldarondo, & F. Mederos (Eds.), *Men who batter: Intervention and prevention strategies in a diverse society* (pp. 3-1–3-20). New York: Civic Research Institute.
- Austin, J., & Dankwort, J. (1999). Standards for batterer programs. *Journal of Interpersonal Violence*, *14*, 152–169.
- Babcock, J., & 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 Publishers.
- Babcock, J., & Robie, C. (in press). Does batterers’ treatment work? A meta-analytic review of domestic violence treatment. *Clinical Psychology Review*.
- Babcock, J., & 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.
- Baumeister, R., Smart, L., & Boden, J. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, *103*, 5–33.
- Belenko, S. (1998). Research on drug courts: A critical review. *National Drug Court Institute Review*, *1*, 1–42.
- Bennett, L. (1998). In defense of batterer-program standards. *Families in Society*, *79*, 93–97.
- Bollen, K., Guilkey, D., & Mroz, T. (1995). Binary outcomes and endogenous explanatory variables: Tests and solutions with an application to the demand for contraceptive use in Tunisia. *Demography*, *32*, 111–131.
- Boruch, R., Snyder, B., & DeMoya, D. (2000). The importance of randomized field trials. *Crime and Delinquency*, *46*, 156–181.
- Budman, S., & Gurman, A. (1988). *Theory and practice of brief therapy*. New York: Guilford.

- Carrillo, R., & Tello, J. (Eds.) (1998). *Family violence and men of color: Healing the wounded male spirit*. New York: Springer.
- Cohen, J. (1994). The Earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Craig, R. J. (1995). Interpersonal psychotherapy and MCMI-III-based assessment. In P. Retzlaff (Ed.), *Tactical psychotherapy of the personality disorders: An MCMI-III-based approach* (pp. 66–88). Boston: Allyn and Bacon.
- Daly, J., & Pelowski, S. (2000). Predictors of dropout among men who batter: A review of studies with implications for research and practice. *Violence and Victims*, *15*, 137–160.
- Daly, J., Powers, T., & Gondolf, E. (2001). Predictors of batterer program attendance. *Journal of Interpersonal Violence*, *16*, 971–991.
- Davis, R., & Taylor, B. (1999). Does batterer treatment reduce violence? A synthesis of the literature. *Women and Criminal Justice*, *10*, 69–93.
- Davis, R., Taylor, B., & Maxwell, C. (1998). *Does batterer treatment reduce violence? A randomized experiment in Brooklyn*. Final report to the National Institute of Justice, Washington, DC.
- Dobash, R. E., & Dobash, R. P. (1992). *Women, violence, and social change*. New York: Routledge.
- Dobash, R. E., & Dobash, R. P. (2000). Evaluating criminal justice interventions for domestic violence. *Crime and Delinquency*, *46*, 252–271.
- Dobash, R. E., Dobash, R. P., Cavanagh, K., & Lewis, R. (2000). *Changing violent men*. Thousand Oaks, CA: Sage Publications.
- Dunford, F. (2000a). Determining program “success”: The importance of employing experimental research designs. *Crime and Delinquency*, *46*, 425–436.
- Dunford, F. (2000b). The San Diego Navy Experiment: An assessment of interventions for men who assault their wives. *Journal of Consulting and Clinical Psychology*, *68*, 468–476.
- Durlak, J., & Lipsey, M. (1991). A practitioner’s guide to meta-analysis. *American Journal of Community Psychology*, *19*, 291–332.
- Dutton, D. (1998). *The abusive personality: Violence and control in intimate relationships*. New York: Guilford.
- Dutton, D., Bodnarchuk, M., Kropp, R., Hart, S., & Ogloff, J. (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., & Kropp, P. (2000). A review of domestic violence risk instruments. *Trauma, Violence and Abuse*, *1*, 171–181.
- Edleson, J. (1996). Controversy and change in batterers’ programs. In J. Edleson, & Z. Eisikovits (Eds.), *Future interventions with battered women and their families* (pp. 154–169). Thousand Oaks, CA: Sage Publications.
- Edleson, J., & Bible, A. (2000). Collaborating for women’s safety: Partnerships between research and practice. In C. Renzetti, J. Edleson, & R. Bergen (Eds.), *Sourcebook on violence against women* (pp. 73–96). Thousand Oaks, CA: Sage Publications.
- Edleson, J., & Syers, M. (1990). The relative effectiveness of group treatments for men who batter. *Social Work Research and Abstracts*, *26*, 10–17.
- Efron, B., & Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association*, *86*, 9–17.
- Elkin, I., Shea, M., & Watkins, J. (1989). The national institute of mental health treatment of depression collaborative research program: General effectiveness of treatments. *Archives of General Psychiatry*, *46*, 971–982.
- Feder, L., & Dugan, L. (2002). A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward experiment. *Justice Quarterly*, *19*, 343–375.
- Feder, L., Jolin, A., & Feyerharm, W. (2000). Lessons from two randomized experiments in criminal justice settings. *Crime and Delinquency*, *46*, 380–400.
- Fein, R., Vossekuil, B., & Holden, G. (1995, July). Threat assessment: An approach to prevent targeted violence. *National Institute of Justice Research in Action*, 1–7.
- Furby, L., Blackshaw, L., & Weinrott, M. (1989). Sex offender recidivism: A review. *Psychological Bulletin*, *105*, 3–30.

- Gamache, D., Edleson, J., & Schock, M. (1988). Coordinated police, judicial and social service response to woman battering: A multi-baseline evaluation across three communities. In: G. Hotaling, D. Finkelhor, J. Kirkpatrick, & M. Straus (Eds.), *Coping with family violence: Research and policy perspectives* (pp. 193–209). Newbury Park, CA: Sage Publications.
- Gendreau, P., Little, T., & Goggin, C. (1996). A Meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 575–607.
- Gladwell, M. (2000). *The Tipping Point: How little things can make a big difference*. New York: Little, Brown.
- Gondolf, E. (1985). *Men who batter: An integrated approach for stopping wife abuse*. Holmes Beach, FL: Learning Publications.
- Gondolf, E. (1987). Changing men who batter: A developmental model of integrated interventions. *Journal of Family Violence*, 2, 345–369.
- Gondolf, E. (1988). Who are these guys? Toward a behavioral typology of batterers. *Violence and Victims*, 3, 187–203.
- Gondolf, E. (1995). Alcohol abuse, wife assault, and power needs. *Social Service Review*, 69, 275–284.
- Gondolf, E. (1997a). Expanding batterer program evaluations. In G. Kaufman, & J. Jasinski (Eds.), *Out of darkness: Contemporary research perspectives on family violence* (pp. 208–218). Thousand Oaks, CA: Sage Publications.
- Gondolf, E. (1997b). Patterns of reassault in batterer programs. *Violence and Victims*, 12, 373–387.
- Gondolf, E. (1998). *The Impact of mandatory court review on batterer program compliance: An evaluation of the Pittsburgh Municipal Court and the Domestic Abuse Counseling Center (DACC)*. Report submitted to Pennsylvania Commission on Crime and Delinquency, Harrisburg, PA.
- Gondolf, E. (1999a). A comparison of reassault rates in four batterer programs: Do court referral, program length and services matter? *Journal of Interpersonal Violence*, 14, 41–61.
- Gondolf, E. (1999b). Characteristics of court-mandated batterers in four cities. *Violence Against Women*, 5, 1277–1293.
- Gondolf, E. (1999c). MCMI results for batterer program participants in four cities: Less “pathological” than expected. *Journal of Family Violence*, 14, 1–17.
- Gondolf, E. (2000a). How batterer program participants avoid reassault. *Violence Against Women*, 6, 1204–1222.
- Gondolf, E. (2000b). Human subject issues in batterer program evaluation. *Journal of Aggression, Maltreatment, and Trauma*, 4(1), 273–297.
- Gondolf, E. (2000c). Mandatory court review and batterer program compliance. *Journal of Interpersonal Violence*, 15, 437–438.
- Gondolf, E. (2000d). Reassault at 30-months after batterer program 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. (2002a). *Batterer intervention systems: Issues, outcomes, and recommendations*. Thousand Oaks, CA: Sage Publications.
- Gondolf, E. (2002b). Service barriers for battered women with male partners in batterer programs. *Journal of Interpersonal Violence*, 17, 217–227.
- Gondolf, E. (in press). *Attendance in specialized batterer counseling for African American men: A clinical trial*. Social Work Research.
- Gondolf, E., & Beeman, A. (2003). Are men who repeatedly reassault distinguished by a special kind of violence? *Violence Against Women*, 9, 1–24.
- Gondolf, E., Chang, Y., & Laporte, R. (1999). Capture–recapture analysis of batterer reassaults: An epidemiological innovation for batterer program evaluation. *Violence and Victims*, 14, 191–202.
- Gondolf, E., Heckert, A., & Kimmel, C. (2002). Non-physical abuse among batterer program participants. *Journal of Family Violence*, 17, 293–315.
- Gondolf, E., & Jones, A. (2001). The program effect of batterer programs in three cities. *Violence and Victims*, 16, 693–704.

- Gondolf, E., & Russell, D. (1986). The case against anger control treatment programs for batterers. *Response*, 9(3), 2–5.
- Gondolf, E., & White, R. (2000). “Consumer” recommendations for batterer programs. *Violence Against Women*, 6, 196–215.
- Gondolf, E., & White, R. (2001). Batterer program participants who repeatedly reassault: Psychopathic tendencies and other disorders. *Journal of Interpersonal Violence*, 16, 361–380.
- Gondolf, E., & Williams, O. (2001). Culturally-focused batterer counseling for African-American men. *Trauma, Violence, and Abuse*, 4, 283–295.
- Gondolf, E., Yllo, K., & Campbell, J. (1997). Collaboration between researchers and advocates. In G. Kaufman, & J. Jasinski (Eds.), *Out of darkness: Contemporary research perspectives on family violence* (pp. 255–261). Thousand Oaks, CA: Sage Publications.
- Gordis, E. (1997, April). Patient–treatment matching—A commentary. *Alcohol Alert (NIAAA Bulletin)*, 36, 3.
- Gottfredson, D., Najaka, S., & Kearley, B. (2003). Effectiveness of drug treatment courts: Evidence from a randomized trial. *Criminology and Public Policy*, 2, 171–196.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: Sage Publications.
- Hamberger, L., Lohr, J., Bonge, D., & Tolin, D. (1996). A large sample empirical typology of male spouse abusers and its relationship to dimensions of abuse. *Violence and Victims*, 11, 277–292.
- Hanson, R. K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers*. User Report 2000-06. Ottawa: Department of the Solicitor General of Canada.
- Heckert, A., & Gondolf, E. (2000a). Assessing assault self-reports by batterer program participants and their partners. *Journal of Family Violence*, 15, 181–197.
- Heckert, A., & Gondolf, E. (2000b). Predictors of underreporting of male violence by batterer program participants and their partners. *Journal of Family Violence*, 15, 423–443.
- Heckert, A., & Gondolf, E. (2000c). The effect of perceptions of sanctions on batterer program outcome. *Journal of Research in Crime and Delinquency*, 37, 369–391.
- Heckert, A., & Gondolf, E. (2001). *Predicting levels of abuse and reassault among batterer program participants*. Final report to the National Institute of Justice, Washington, DC.
- Heckert, A., Matula, D., & Gondolf, E. (2000). Women’s accounts of domestic violence. *Women and Criminal Justice*, 12, 95–120.
- Heckman, J. (1979). Sample selection as a specification error. *Econometrica*, 47, 153–161.
- Holtzworth-Munroe, A., Meehan, J., Herron, K., Rehman, U., & Stuart, G. (2001). Testing the Holtzworth-Munroe and Stuart (1994) batterer typology. *Journal of Consulting and Clinical Psychology*, 69, 1000–1020.
- Holtzworth-Munroe, A., & Stuart, G. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin*, 116, 476–497.
- Hubbard, R., Marsden, M., Rachal, J., Harwood, H., Cavanaugh, E., & Ginzburg, H. (1989). *Drug abuse treatment, A national survey of effectiveness*. Chapel Hill, NC: University of North Carolina Press.
- Jacobson, N., & Gottman, J. (1998). *When men batter women: New insights into ending abusive relationships*. New York: Simon and Shuster.
- Jennings, J. (1990). Preventing relapse versus “stopping” domestic violence: Do we expect too much too soon from battering men? *Journal of Family Violence*, 5, 43–60.
- Johnson, H. (2000). The role of alcohol in male partners’ assaults on wives. *Journal of Drug Issues*, 30, 725–740.
- Jones, A. (1998). Sample attrition in panel data from an evaluation of four batterer intervention programs. *Paper presented at the Program Evaluation and Family Violence Research Conference, Durham, NH*.
- Jones, A. (2000). The cost of batterer treatment: How much? Who pays? *Journal of Interpersonal Violence*, 15, 566–586.
- Jones, A., D’Agostino, R., Gondolf, E., & Heckert, A. (in press). Assessing the effect of batterer program completion on reassault using propensity scores. *Journal of Interpersonal Violence*.
- Jones, A., & Gondolf, E. (2001). Time-varying risk factors for reassault by batterer program participants. *Journal of Family Violence*, 16, 345–359.

- Jones, A., & Scharfstein, D. (2001). Use of the odds ratio as a measure of effect size. *Manuscript submitted for publication*.
- Kassebaum, G., & Ward, D. (1991). Analysis, reanalysis and meta-analysis of correctional treatment effectiveness: Is the question what works or who works? *Sociological Practice Review*, 2, 159–168.
- Kivel, P. (1992). *Men's work: How to stop the violence that tears our lives apart*. Center City, MN: Hazelden.
- Laing, K., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Langford, D. (1996). Predicting unpredictability: A model of women's processes of predicting battering men's violence. *Scholarly Inquiry for Nursing practice*, 10, 371–385.
- Levesque, D., & Gelles, R. (1998). Does treatment reduce violence recidivism in men who batter? Meta-analytic evaluation of treatment outcome research. (Paper presented at). "Program Evaluation and Family Violence Research" Conference, Durham, New Hampshire (pp. 26–29).
- Loza, W., & Loza-Fanous, A. (1999a). Anger and prediction of violent and nonviolent offenders' recidivism. *Journal of Interpersonal Violence*, 14, 1014–1029.
- Mann, C. (1990). Meta-analysis in the breech. *Science*, 249, 476–480.
- Marshall, W., & Pithers, W. (1994). A reconsideration of treatment outcome with sex offenders. *Criminal Justice and Behavior*, 21, 10–28.
- Marshall, W., & Serran, G. (2000). Improving the effectiveness of sexual offender treatment. *Trauma, Violence, and Abuse*, 1, 203–222.
- Maxwell, C., Garner, J., & Fagan, J. (2001, June). The effects of arrest on intimate partner violence: New evidence from the Spouse Assault Replication Program. *National Institute of Justice Research in Brief*, 1–15.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, 71, 173–180.
- Mederos, F. (1999). Batterer intervention programs: The past, and future prospects. In M. Shepard, & E. Pence (Eds.), *Coordinating community responses to domestic violence: Lessons from Duluth and beyond* (pp. 127–150). Thousand Oaks, CA: Sage Publications.
- Millon, T. (1994). *Millon Clinical Multiaxial Inventory-III manual*. Minneapolis, MN: National Computer Systems.
- Monahan, J., Steadman, J., Sliver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of mental disorder and violence*. Chicago: University of Chicago Press.
- Mulvey, E., & Lidz, C. (1985). Conditional prediction: A model for research on dangerousness to others in a new era. *International Journal of Law and Psychiatry*, 18, 129–143.
- Murphy, C., & Dienemann, J. (1999). Informing the research agenda on domestic abuser intervention through practitioner-researcher dialogues. *Journal of Interpersonal Violence*, 14, 1314–1327.
- Murphy, C., Musser, P., & Maton, K. (1998). Coordinated community intervention for domestic abusers. *Journal of Family Violence*, 13, 263–285.
- National Opinion Research Center, K. (1996). *The national treatment improvement evaluation study: The persistent effects of substance abuse treatment—one year later*. Washington, DC: Substance Abuse and Mental Health Services Administration, U.S. Department of Health and Human Services.
- National Violence Against Women Prevention Research Center (2001). *Fostering collaborations to prevent violence against women: Integrating findings from practitioner and research focus groups*. Charleston, SC: Author.
- O'Leary, K. (2002). Conjoint therapy for partners who engage in physically aggressive behavior: Rationale and research. *Journal of Aggression, Maltreatment, and Trauma*, 5(2), 145–164.
- Pawson, R., & Tilly, N. (1997). *Realistic evaluation*. Thousand Oaks, CA: Sage Publications.
- Pence, E., & Paymar, M. (1993). *Education groups for men who batter: The Duluth model*. New York: Springer.
- Pence, E., & Shepard, M. (1999). Developing a coordinated community response. In M. Shepard, & E. Pence (Eds.), *Coordinating community responses to domestic violence: Lessons from Duluth and beyond* (pp. 3–25). Thousand Oaks, CA: Sage Publications.
- Project MATCH Research Group (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcome. *Journal of Studies on Alcohol*, 58, 7–29.

- Quinsey, V., Harris, G., Rice, M., & Lalumiere, M. (1993). Assessing treatment efficacy in outcome studies of sex offenders. *Journal of Interpersonal Violence*, 8, 512–523.
- Raab, S. (2000). Men explode: Have you ever hit a woman? *Esquire Magazine*, 134(3), 224–230.
- Rice, M. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist*, 52, 414–423.
- Roehl, J., & Guertin, K. (2000). Intimate partner violence: The current use of risk assessments in sentencing offenders. *Justice System Journal*, 21, 171–197.
- Rondeau, G., Brodeur, N., Brochu, S., & Lemire, G. (2001). Dropout and completion of treatment among spouse abusers. *Violence and Victims*, 16, 127–143.
- Rosenfeld, B. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review*, 12, 205–226.
- Russell, M. (1995). *Confronting abusive beliefs: Group treatment for abusive men*. Thousand Oaks, CA: Sage Publications.
- Saunders, D. (1996). Feminist–cognitive–behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and Victims*, 11, 393–414.
- Saunders, D., & Hamill, R. (in press). *Offender interventions to end violence against women: A research synthesis for practitioners*. Washington, DC: National Institute of Justice, U.S. Department of Justice.
- Schare, M. (1992). Needed directions in the psychotherapeutic treatment of drunk drivers. (Paper presented at). *International Perspectives on Crime, Justice, and Public Order, St. Petersburg, Russia*.
- Selzer, M. (1971). The Michigan Alcoholism Screening Test (MAST): The quest for a new diagnostic instrument. *American Journal of Psychiatry*, 3, 176–181.
- Skeem, J., & Mulvey, E. (2001). Psychopathy and community violence among civil psychiatric patients: Results from the MacArthur Violence Risk Assessment Study. *Journal of Consulting and Clinical Psychology*, 69, 358–374.
- Smith, P., & Smith, J. (1999). Beyond the measurement trap. *Psychology of Woman Quarterly*, 23, 177–194.
- Steadman, H. (1982). A situational approach to violence. *International Journal of Law and Psychiatry*, 5, 171–186.
- Stordeur, R., & Stille, R. (1989). *Ending men's violence against their partners*. Thousand Oaks, CA: Sage Publications.
- Stosny, S. (1995). *Treating attachment abuse: A compassion approach*. New York: Springer.
- Straus, M. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) scales. *Journal of Marriage and the Family*, 41, 75–88.
- Sutton, S. (2001). Back to the drawing board? A review of applications of the transtheoretical model to substance use. *Addiction*, 96, 175–187.
- Taubman, S. (1986). Beyond the bravado: Sex roles and the exploitive male. *Social Work*, 31, 12–18.
- Tolman, R., & Bennett, L. (1990). A review of quantitative research on men who batter. *Journal of Interpersonal Violence*, 5, 87–118.
- Van Voohis, P., Cullen, F., & Applegate, B. (1995). Evaluating interventions with violent offenders: A guide for practitioners and policy makers. *Federal Probation*, 56(3), 17–27.
- Walters, G. (2000). Disposed to aggress? In search of the violence-prone personality. *Aggression and Violent Behavior*, 5, 177–190.
- Weisz, A., Tolman, R., & Saunders, D. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence*, 15, 75–90.
- White, R., & Gondolf, E. (2000). Implications of personality profiles for batterers treatment. *Journal of Interpersonal Violence*, 15, 467–488.
- Williams, O. (1994). Group work with African-American men who batter: Toward more ethnically-sensitive practice. *Journal of Comparative Family Studies*, 25(1), 91–103.
- Williams, O. (1995). Treatment for African-American men who batter. *CURA Reporter: Bulletin of the Center for Urban and Regional Affairs, University of Minnesota*, vol. 25(3) (pp. 12–16). Minneapolis, MN: University of Minnesota.
- Wilson, D., & Lipsey, J. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, 6, 413–429.