



Controversies and Recent Studies of Batterer Intervention Program Effectiveness

Larry Bennett and Oliver Williams

Batterer intervention programs (BIPs) are designed for men arrested for domestic violence and for men who would be arrested if their actions were public. These programs usually consist of educational classes or treatment groups, but may include other intervention elements such as extensive evaluation, individual counseling, or case management. Because 80% of batterers are referred by the criminal justice system, one set of implicit goals for BIPs includes justice and accountability (Healy, Smith & O'Sullivan, 1998), goals that have not been adequately recognized in evaluations of BIPs. Another goal of BIPs is victim safety. Most standards for BIPs specify that service providers consider victim safety implications when implementing interventions such as contacting victims for information about the batterer (Austin & Dankwort, 1997; 1999). A final goal for BIPs is rehabilitation and behavioral changes such as skill building, attitude change, and emotional development.

The details of conducting batterer intervention programs are readily available (e.g. Edleson & Tolman, 1992; Pence & Paymar, 1993; Russell & Frohberg, 1995; Sonkin & Durphy, 1997; Stordeur & Stille, 1989). The purpose of this paper is to look not at what batterer programs do, but rather at the effectiveness of these programs. Knowledge about batterer program effectiveness is important for several reasons. Increasingly, courts are referring men convicted of domestic abuse to batterer intervention programs, suggesting a certain level of public confidence in the effectiveness of these programs. Is that confidence justified? Second, victims of domestic violence often want to remain in a relationship with their partner, and are looking for help in changing his violent and controlling behavior.

Since a batterer seeking counseling is one of the strongest predictors that a woman will return to her batterer (Gondolf, 1988), advocates are justifiably concerned that batterer programs not hold out a promise of hope which may become a vehicle for injury. Third, people who work with batterers are interested in outcomes so they can improve the level of program effectiveness; for these people, the concern is less *whether* batterer programs work, but *how* they work, *for whom* do they work best, and *which elements* of the program are most important.

Issues in BIP Evaluation

The questions we want answered are: (1) Are batterers held accountable for their crime (or, has justice been served)? (2) Are victims safe? And, (3) Has the batterer changed his attitudes and behavior? The question of whether justice is being served is not easily answered by science. Batterer programs often identify accountability as a key theme in their work, but BIPs have neither the mandate nor the resources to hold men accountable for their actions; for that task, BIPs are but a link in the chain. Victim safety is the "gold standard" for batterer intervention programs, the primary criterion by which program effectiveness will be judged. Recidivism is an indicator that victim safety has been breached. The third question, behavioral change, taps the skill-building and attitude modification focus of many batterer programs. Do batterers acquire skills and change their beliefs about women and the acceptability of violence as a result of batterer programs? This is an important question for program evaluation. However, attitude and skills changes must be viewed

as intermediate factors that service the most fundamental goal: safety of past, present, and future victims.

How is victim safety evaluated? One way is to ask victims whether they are safe during and after a batterer's participation in a BIP. Victims are the best reporters of current abuse, and they are often reliable predictors of future abuse (Weisz, Tolman, & Saunders, 2000). Contacting victims may be problematic, however. Victims of domestic violence often change residence and phone, and their relatives (if known) are understandably hesitant about releasing information to strangers. Victims and batterers are often separated, and the batterer may reside with another partner during and after the batterer program. In one large study of batterer programs in four U.S. cities, 21 percent of the batterers in the study had new partners by the 30-month follow-up (Gondolf, 1998). While victim report is the most reliable indicator of repeat violence, other indicators of victim safety may be employed in studies of batterer program effectiveness.

The most common substitute for victim report is official records of the police or court, such as restraining orders and records of arrest or conviction. While official records may be a more convenient indicator of program success, use of official records as outcome is complicated by the fact that those batterers most likely to not attend or drop out of batterer programs are the ones most likely to be re-arrested. Some studies include only those batterers who complete the program, but program completers are disproportionately white, middle class, employed, and married, in part because the facilitators of batterer programs are also disproportionately white, middle class, employed, and married (Williams & Becker, 1994). Moreover, the chances of being re-arrested for domestic abuse is only a fraction of the chances of re-abuse, so use of official records as an outcome under-estimates recidivism and over-estimates the effects of the batterer program. For example, one careful study found that the proportion of arrest to victim-reported abuse was 1 in 35; that is, for every reported arrest, there were 35 assaultive actions (Dutton, et al., 1997).

Use of physical violence as an indication of recidivism, while overlooking non-physical forms of abuse and control, also complicates BIP outcome studies. Some researchers argue that non-physical abuse and control is a qualitatively different form of behavior than physical abuse, with different risk factors (O'Leary, 1993). Nevertheless, much of the content of contemporary batterer intervention programs is focused on learning non-controlling behavior (Healy, Smith, & O'Sullivan, 1998). A long-standing suspicion of batterer intervention is that men may learn to avoid physical abuse by substituting more economical and legal forms of control such as intimidation, isolation, and surveillance. Abusive men may also punish both their victims and their children through protracted child custody and visitation cases. Consequently, ignoring non-physical abuse over-estimates the effectiveness of batterer programs.

Evaluating the effectiveness of batterer intervention programs is further complicated by the frequent co-occurrence of other problems, most notably unemployment, substance abuse, and mental disorders. These co-occurring risk factors are not usually viewed as the cause of violence, but their co-existence makes intervention more difficult and outcomes more negative. Men with these co-occurring problems are far more likely to drop out of a BIP. Some evaluations have attempted to control for substance abuse and mental disorders by excluding these men from their sample (e.g. Dunford, 2000). Unfortunately, since these co-occurring problems are so common in day-to-day BIPs, the validity of a study which excludes "dual problem" batterers is seriously compromised. Attrition is a substantial obstacle for BIPs, with many studies reporting that less than half the referred batterers complete the program. Dropping out of BIPs is further reinforced by the lack of sanction in many communities for failing to attend the program. In fact, Frank (1999) has suggested that the most important "outcome indicator" is not individual behavior or recidivism, but rather community behavior: specifically, the community response to batterer non-compliance.

Confidence in BIP Evaluation Studies

Researchers often divide studies into three categories: non-experimental, quasi-experimental, and experimental. These categories reflect the confidence in the results of an evaluation, with highest confidence accorded to experimental evaluations, and lowest to non-experimental evaluations. Non-experimental studies cannot attribute the outcome to the BIP. With no group of batterers who were not in the program to whom we can compare the program participants, we are unable to differentiate the effect of the BIP from the effect of the many other factors which may prevent further violence. A non-experimental BIP evaluation typically measures recidivism by either re-arrest or victim report during a period following the intervention. If the results of such a study report 70% of the men completing the BIP are non-violent during the 12 months following program completion B a typical report B it is not appropriate to argue that the BIP prevented violence in seven of ten participants. Other treatment, medication, outside events, and any number of unmeasured variables are likely to effect recidivism, and there is no way to tell which factors are at work. So-called “stake in conformity” variables such as marital status, employment, and history of arrest often predict both who completes batterer programs and whether they re-offend (Feder & Forde, 2000; Toby, 1957). Despite these limitations, non-experimental studies are essential pre-cursors to evaluations using quasi-experimental and experimental methods.

Quasi-experimental designs are more rigorous than non-experimental designs, and use some form of comparison group to control for the effects of other factors. One form of comparison group is that group which is naturally formed by men who should be in the BIP but are not, either because they never attended, were subsequently excluded, or else dropped out after they started the program. We can have only limited confidence in these designs using program dropouts because the characteristics of men who drop out of BIPs often differ from the characteristics of men who complete the program. In fact, dropout characteristics are similar to charac-

teristics of those men most likely to re-offend: unemployed, young, substance abuser, and not in a stable relationship. When program completers are subsequently compared to program dropouts, it “lowers the bar” and makes the program appear more effective, because the men who complete the program are being compared to more marginalized men. One way of compensating for these differences is to use statistics to artificially control for any observed differences between completers and dropouts.

Men who do not get the BIP, but who get something else instead, such as probation or alternative service form a second type of comparison group. One of the problems with this kind of study is that men who are sentenced to BIPs are often substantially different than men sentenced to the alternative condition. To adjust for these differences, evaluators again use statistics to control for differences between the two groups. However, there may be differences between the BIP men and the comparison group men that are not measured, and there is no way to account statistically for these unseen factors. Moreover, some of the sentencing alternatives provided to men in the comparison group may have violence-reducing positive effects. For example, assigning men to community service such as work at a nursing home may increase their sense of empathy. The end result is that the comparison group gets “treated” and the effects of the BIP are minimized.

The best solution to the problem of dissimilarity between BIP participants and those men getting other interventions is random assignment to a BIP or a control group. Men in the control group get either no service (which is unlikely) or some form of usual and customary intervention, such as probation. Batterers may also be assigned to different BIP conditions such as treatment intensity (Edleson and Syers, 1990) or intervention model (Edleson & Syers, 1990; Saunders, 1996). Random assignment increases the chances that all the unmeasured factors which make one batterer different than another are as likely to be in the BIP condition as the control condition, with the end result being a canceling out of those effects. On paper, this is the optimal

method to evaluate BIPs, but random assignment is a very difficult procedure. If randomization is done at the point of sentencing, the judge, prosecutor, and defense must all agree to it. Judges often feel compelled to break with random assignment due to the characteristics of a certain case, usually to refer the batterer to a BIP rather than to the alternative condition. Prosecutors also may object to the batterer not being in a BIP because they view the BIP both as a deterrent from future crime and as punishment for past crime. Despite the difficulty with experimental designs, a number of experimental studies have been completed, and four such evaluations of batterer intervention programs which used no-treatment or customary treatment control groups are now available: the Ontario experiment (Palmer, Brown, & Barerra, 1992), the San Diego Navy experiment (Dunford, 2000), the Broward experiment (Feder & Forde, 2000), and the Brooklyn experiment (Taylor, Davis, & Maxwell, 2001).

Results of BIP Evaluations

The main questions to be addressed are: (1) Are batterer intervention programs effective when compared to customary practice (usually probation)? and (2) Are certain approaches to batterer intervention programming more effective than other approaches? Our conclusions will be these: (1) Batterers programs as currently configured have modest but positive effects on violence prevention, and (2) there is little evidence at present supporting the effectiveness of one BIP approach over another.

Table 1 (page 5) summarizes the four experimental studies from which we can best draw conclusions about the first question: are batterer intervention programs effective when compared to customary practice? A short description of these randomized control group evaluations is in the Appendix of this paper, along with other important experimental and quasi-experimental studies which help us answer our second question. Readers who want more detailed analyses of the empirical evaluations of batterer programs are referred to the growing body of outcome review literature (e.g. Babcock & LaTaillade, 2000; Davis & Taylor, 2000; Eisikovits & Edleson,

1989; Gondolf, 1987; Rosenfeld, 1992; Tolman & Bennett, 1990; Tolman & Edleson, 1995)

As we see in Table 1, two of the four experiments (Dunford, 2000; Feder & Forde, 2000) found no difference in recidivism for men in the batterer program and men in the control condition. The other two experiments (Palmer, et al., 1992; Taylor, Davis, & Maxwell, 2001) found small but significant reductions in recidivism for men in batterer programs. While it is beyond the scope of this paper to provide a detailed analysis of these experiments, they make such an important contribution to our understanding of batterers program that they merit the short descriptions that follow.

Ontario Experiment

The first experiment comparing batterer programs to a control group was conducted by Palmer and her colleagues in Hamilton, Ontario. Palmer, Brown, and Barrera (1992) studied 59 men convicted of wife abuse, placed on probation, and randomly assigned to either a 10-week batterer program at a local family service agency, or to probation with no batterer program. The intervention was characterized by the researchers as psycho-educational and client-centered. Seventy percent of the BIP participants completed their program, and 87% attended at least half the sessions. A year after the program ended, all subjects and partners were mailed questionnaires followed by phone calls, but the response rate was low. Police records were searched for complaints or arrests. Three of the 30 (10%) men assigned to the batterer program re-offended, according to police records, compared to eight of 26 (31%) men receiving probation only. Most criticism of this study focuses on the small number of participants. Had the results not been favorable to batterer programs, BIP proponents would have also pointed out that 15 hours of group would not meet the program standards in most states which have such standards (Austin & Dankwort, 1997; 1999). Despite concerns about sample size and intervention dose, the Ontario study is a solid addition to the findings of non-experimental and quasi-experimental evaluations. The study provides support for the

modest effectiveness of short-term batterer intervention programs.

Navy Experiment

Dunford (2000) reports the results of an experiment at the Navy base in San Diego where 861 men who assaulted their wives were randomly assigned to one of four conditions: (a) six months of weekly cognitive-behavioral treatment, followed by six months of monthly groups; (b) six months of group for couples, followed by six months of monthly group; (c) a rigorous monitoring and case management program similar to probation, or (d) safety planning, similar to the work of victim advocates, which serves as a control group. Seventy percent of the men completed their program. In Table 1, we consider only (a) the BIP and (d) the control group. Standard practice in batterer intervention excludes groups for couples as a threat to victim safety, and in fact, two thirds of the female members of the couples in (b) were not present during the couples group, possibly voting with their feet on the popularity among victims of the couples' model. Dunford found no significant differences between the four groups. Were the experiment generalizable to other batterers programs, we would conclude that batterers programs had no significant effect on domestic abuse. The problem with accepting the Navy experiment is the characteristics of its participants. Excluded from the Navy experiment, either by design or by circumstance, were the following: substance abusers, men with mental disorders, men with prior criminal records, unmarried men, and unemployed men. Furthermore, the group was offered by the men's employer, at their place of employment. In fact, most of the men who would be seen in typical BIPs were excluded from this study. The Navy experiment, while questionable as an indicator of batterer program effectiveness, is nevertheless useful as an indicator of coordinated community intervention. The overall recidivism rate was 30% by spouse report and 4% by arrest. These figures compare very favorably with other interventions. What we can conclude from the Navy experiment is this: If communities take a proactive response to domestic violence, including assertive

probation work, sanctions for non-compliance, victim safety monitoring, and batterer intervention programs, they will probably reduce the incidence of repeat violence.

Broward Experiment

Feder and Forde (2000) studied all 404 male defendants convicted of misdemeanor domestic violence in Broward County Florida (Fort Lauderdale) over a five month period. Men were randomly assigned to either probation and six months of a Duluth model BIP (Pence & Paymar, 1993) or probation only. Researchers collected information on minor and severe abuse, violations of probation and re-arrests using offender self-reports, victim reports, and official measures. Ninety-five percent of the men assigned to the BIP attended at least 20 of 26 meetings, a rather remarkable figure when compared to the average BIP attrition rate of 50%. Since less than a third of victims could be interviewed at follow-up, these results are not included in Table 1. At 12-month follow-up, there were no differences between the BIP participants and regular probationers on measures of attitude toward women, beliefs about wife-beating, attitudes toward treating domestic violence as a crime, beliefs about the female partner's responsibility for the violence, estimated chance of hitting partner in the next year, and victim or official report of recidivism. One of the key findings of the Broward experiment was further support for the *stake in conformity* hypothesis: men most likely to re-offend are those who have the least to lose, as measured by education, marital status, home ownership, employment, income, and length of residency. This finding is robust over a number of BIP studies and presents one of the most formidable obstacles to effective batterer intervention programs, as well as evaluating those programs.

Brooklyn Experiment

Taylor, Davis, and Maxwell (2001) report the findings of 376 men convicted of misdemeanor domestic violence and randomly assigned to 40 hours of a Duluth model BIP or 40 hours of community service. Victim reports and official records

were used to track differences at six-month and 12-month follow-up. At follow-up with partners, BIP participants were more likely than controls to have been abusive, but the difference was not significant. Using criminal justice records, BIP participants were 50% less likely to have re-offended at both six-month and 12-month follow-ups. However, enthusiasm for this result is tempered by the fact that judge, prosecutor, and defendant had to agree on the man's referral to the BIP, a process which effectively screened out men with low motivation.

Considering these four experiments, along with a growing body of quasi-experimental and non-experimental studies, we conclude that the effect of BIPs is modest, but nevertheless significant. By significant, we do not necessarily mean statistically significant, but rather practically significant. Asking whether batterer programs are more effective than probation alone is asking the wrong question because batterer programs were never designed to be used instead of probation. Augmenting the influence of probation and providing an additional vehicle for accountability is one of the goals of batterer programs. If there were no statistically positive effects for batterer programs, which is clearly not the case according to the research, then we could rightly say these programs were not effective. The best statement we can make at this time is that BIPs add a small but important effect to overall violence prevention.

Are Some Approaches Better Than Others?

If BIPs have a small effect - or even if there were no effect - we would want to know if there are characteristics of programs which may lead to greater effectiveness, and whether some approaches are more effective than others. The program parameters most often discussed are theoretical foundation, program length and structure, and the extent to which the program attends to co-existing problems such as substance abuse or mental/personality disorders. The questions we might ask of the research are these: (1) Is there a "model" of BIPs which is more effective than others? (2) Are structured, psycho-education programs more effective

than unstructured, process-oriented programs? (3) Are longer programs more effective than shorter programs? and, (4) Are integrated, mental health-focused programs more effective than programs which do not attend to co-occurring problems?

At present we have only a few studies which address these issues. Several studies have compared models, including psycho-educational and couples groups (Dunford, 2000; Brannen & Rubin, 1996; O'Leary, Heyman, & Neidig, 1999), psycho-educational and self-help groups (Edleson & Syers, 1990, 1991), and psychodynamic and cognitive-behavioral groups (Saunders, 1996). The results of studies to date suggest no significant differences in outcome between models, with the exception that men with high levels of dependency may do better in process-oriented groups, while generally violent, antisocial batterers may do better in cognitive-behavioral groups (Saunders, 1996). However, there may be reasons other than evaluated effectiveness to use or not use a particular model. Couple counseling is proscribed in several states, for example, and anger management as a solo model is also proscribed by some state standards. In fact, there may be some empirical support for the latter proscription. Babcock and La Taillade (2000), comparing effect sizes across models, find an average effect size of 0.44 for Duluth-type psycho-educational programs, and an average effect size of 0.14 for anger management programs. In the language of effect sizes, 0.2 is a small effect, 0.5 is moderate, and 0.8 is considered a large effect. This may be an unfair comparison, however, because Duluth-type programs are more likely to be part of a community intervention program, and consequently more likely to benefit from the additive effects of arrest and prosecution, assertive sanctions for non-compliance, victim advocacy and counseling, as well as a batterer program (Murphy, Musser, & Maton, 1998).

Studies of program length and program structure yield similar outcomes of no-difference. The four sites in Gondolf's (1999) multi-site study varied in length from three to nine months, but at 15-month follow-up, there were no significant differences between the four programs in re-assault, threats, or

victim quality of life. Edleson and Syer's (1990) randomly assigned batterers to a more intense condition (32 sessions over 16 weeks) and less intense condition (12 sessions over 12 weeks) but at six-month follow-up, there were no differences in victim-reported re-assault. We conclude that, at present, there is little support for an argument that longer programs are more effective than shorter programs. There may be other reasons to have longer term programs, however. To the extent a goal for batterer programs involves justice and accountability, maintaining the batterer in a regular program for a longer time may promote these goals. Another issue impacting program length may be the influence of third parties. Some third parties, such as an insurance company, may want shorter programs, while other third parties, such as a state agency, may want longer programs. Another reason for longer programs is a possible deterrence effect of batterer programs: while he's in the program, he is more vigilant. However, recent evidence suggests most batterers who re-offend do so within six months of their admittance to the program (Gondolf, 1999a). However, even if batterers re-offend in the last six months, there still may be a safety advantage to having the batterer out of the house for a predictable three hours a week.

There is limited evaluation of program structuring. Edleson and Syers (1990) compared more structured and less structured groups, and found a slight but non-significant effect favoring more structure. Saunders (1996) found that less structured groups appeared to be more effective with dependent men, while more structured groups were more effective with antisocial men. The effect of structure may have less to do with the structure of the program itself than with the structure of the system in which the program operates. There is emerging evidence that coordinated community efforts in which the batterer program plays a necessary but not sufficient role in violence prevention are more effective than situations in which the batterer program is viewed as the singular intervention for men who batter (Babcock & Steiner, 1999; Frank, 1999; Healy, Smith & O'Sullivan, 1998; Murphy, Musser, &

Maton, 1998; Syers & Edleson, 1992).

The co-occurrence of domestic abuse with mental disorders, personality disorders, and substance abuse has been amply documented (e.g. Dutton & Starzomski, 1993; Gondolf, 1999b; Hastings & Hamberger, 1988; Holtzworth-Munroe & Stuart, 1994; Leonard & Jacob, 1988; Murphy, Meyer, & O'Leary, 1993). Men in batterers programs are more likely to have these conditions than either men in the general population or batterers who are not referred to BIPs. Many professionals view battering as, if not a symptom of substance abuse or mental disorders, then at least a confounding factor which impedes the opportunity to learn non-violent behavior. While there is little evidence that substance abuse or mental disorders cause a man to be violent who would not otherwise be violent, the fact that half the men in batterers groups have diagnosable disorders suggest that ethical BIPs should screen for these problems. Gondolf (1999a) found some evidence that the program, which explicitly attended to mental health and substance abuse concerns, had the lowest rate of severe assault recidivism. On the other hand, evaluation of Seattle's coordinated community response found that batterers completing substance abuse treatment were as likely to recidivate as batterers not completing substance abuse treatment (Babcock & Steiner, 1999). Beyond the few caveats from the studies described in this paper, there is little evidence supporting the belief that batterers programs should attend to mental health and substance abuse issues beyond the screening and referral which is currently the standard of practice.

One additional problem with the evaluation literature on batterer programs is the lack of information on culturally competent practice. In many ways, this deficit reflects the current development of the field. One national study of batterer programs found that most are deliberately colorblind, and choose to not address the realities or concerns of men of color (Williams & Becker, 1994). This finding suggests one possible reason for ineffective batterer programs. Comparing men who complete treatment in either racially mixed or African-American groups, Williams (1995) found race to be a significant influence on

trust, comfort, willingness to discuss critical subjects, and participation in treatment. Men in the African-American groups felt more positive about their experiences and more willing to discuss issues associated with race that they considered as influences on their behavior. In the Pittsburgh setting of the Multi-site study, Gondolf and Williams (2001) report that only 52% of the African-American men in the batterer program completed the program, compared to 82% of white men. African-American men in the Pittsburgh program were twice as likely as whites to be re-arrested (13% v. 5%), but were less likely to re-offend as reported by their partners (32% v. 39%). These limited findings suggest the need for a concerted focus on culturally focused batterer programs for African-American men.

Application of Findings to Practice

Considering the information presented above, we offer the following summary statements about the effectiveness of batterer intervention programs, and how research might be applied to practice. These statements should be taken as hypotheses generated from research and practice, not as facts.

(1) BIPs have a small but significant effect. Batterer programs are not *treatments* in the medical or therapeutic sense, so it is not surprising that their effect is small. Batterer programs are critical elements in an overall violence prevention effort. The effect of any of the elements in this effort—education, arrest, prosecution, probation, victim services, adjunct services, and BIP—is diminished by the removal of any of the other efforts. The most effective reduction in partner violence will occur in those communities with the strongest combination of coordinated, accountable elements. The challenge to BIP practitioners is to make sure their practice extends beyond the level of the individual to the level of the community. Practitioners should work to educate and support all elements of a coordinated community response.

(2) BIPs are more effective for some men than others. Whether the effect is analyzed by a

man's stake in conformity (education, employment, relationship commitment, community bonding), mental status (the effects of personality disorder, mental disorder, substance abuse disorder), or cultural congruity (the more group facilitators share culture and language with the participants, the greater the stake in the group), one in four men referred to a BIP will account for most of the repeat violence and most of the serious injury within a batterer program. Since the batterer program alone will not effectively reduce his potential for violence, the batterer program's best role for these men is to hold them in program as long as possible, increasing the time a battered woman may need to get herself into a safer position. Although longer term (26 to 52 week) programs may not be more effective in rehabilitating these batterers, they may serve a more useful function as agents of justice, accountability, and victim safety than shorter term programs.

(3) Assessment must occur on an ongoing basis. Most re-offense occurs early, usually within six months of initial program intake. Assessment and accountability must be on-going, not something which is done only at program intake and follow-up. Ongoing assessments should include both battering and substance abuse.

(4) Encourage experimentation and program development. No program approaches have shown themselves to be superior to other approaches, so standards that mandate specific BIP models cannot be based on effectiveness research. Within the boundary of safety and accountable practice, developing effective programs is more likely under conditions of supervised experimentation. Due to concerns about safety, batterer programs must not only hold their participants accountable, but BIPs must hold themselves accountable as well. The safe way to engage in experimentation to boost program effectiveness is to work closely with criminal justice authorities, a local victim services agency, and victim advocates.

(5) Evaluate outcomes. Evaluation is one mechanism of accountability. Programs which routinely evaluate what they do—and its effectiveness—are likely safer than programs which do not

conduct routine evaluations. Since the hope that a batterer will be helped is a major force for women returning to their abuser, BIPs are responsible to monitor their own capacity for violence prevention. A batterers program alone is not enough.

Distribution Rights: This Applied Research paper and In Brief may be reprinted in its entirety or excerpted with proper acknowledgement to the author(s) and VAWnet (www.vawnet.org), but may not be altered or sold for profit.

Suggested Citation: Bennett, L. and Williams, O. (2001, August). *Controversies and Recent Studies of Batterer Intervention Program Effectiveness*. Harrisburg, PA: VAWnet, a project of the National Resource Center on Domestic Violence/ Pennsylvania Coalition Against Domestic Violence. Retrieved month/day/year, from: <http://www.vawnet.org>

Authors of this document:

Larry Bennett, Ph.D.
University of Illinois at Chicago
Jane Addams College of Social Work

Oliver Williams, Ph.D.
University of Minnesota
School of Social Work
August 2001

Consultant:

Nancy Kreidman
Executive Director
The Domestic Violence Clearinghouse
and Legal Hotline
dvclh@stoptheviolence.org

References

Austin, J. & Dankwort, J. (1997). A review of standards for batterer intervention programs. *VAWNet: Violence Against Women Online Resources*. Retrieved from <http://www.vaw.umn.edu/Vawnet/standard.htm>

Austin, J. & Dankwort, J. (1999). Standards for batterer programs: A review and analysis. *Journal of Interpersonal Violence, 14*, 152-168.

Babcock, J.C., & LaTailedde, J.J. (2000). Evaluating interventions for men who batter. In J.P. Vincent and E.N. Jouriles (Eds.) *Domestic violence: Guidelines for research informed practice*. Philadelphia: Jessica Kingsley Publishers.

Babcock, J.C. & Steiner, R. (1999). The relationship between treatment, incarceration, and recidivism of battering: A program evaluation of Seattle's coordinated community response to domestic violence. *Journal of Family Psychology, 13*, 46-59

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

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

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

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

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

Dutton, D.G., Bodnarchuk, M., Kropp, R., Hart, S.D., & Ogloff, J.R.P. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. *International Journal of Offender Therapy and Comparative Criminology, 41*, 9-23.

Dutton D.G. & Starzomzki, A.J. (1993). Borderline personality in perpetrators of psychological and physical abuse. *Violence and Victims, 8*, 327-337.

- Edleson, J.L., & Syers, M. (1990). The relative effectiveness of group treatment for men who batter. *Social Work Research and Abstracts*, 26(2) 10-17.
- Edleson, J.L. & Syers, M. (1991). The effects of group treatment for men who batter: An 18 month follow-up study. *Research in Social Work Practice*, 1, 227-243.
- Edleson, J.L. & Tolman, R.M. (1992). *Intervention for men who batter: An ecological approach*. Newbury Park, CA: Sage.
- Eisikovits, Z.C. & Edleson, J.L. (1989). Intervening with men who batter: A critical review of the literature. *Social Service Review*, 63, 384-414.
- Feder, L., & Forde, D.R. (2000). *A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward experiment*. National Institute of Justice.
- Frank, P.B. (1999, July). *Measuring the system, not individuals*. Paper presented at the 6th International Family Violence Research Conference, Durham NH.
- Gondolf, E.W. (1987). Evaluating programs for men who batter: Problems and prospects. *Journal of Family Violence*, 2, 95-108.
- Gondolf, E.W. (1988). The effect of batterer counseling on shelter outcome. *Journal of Interpersonal Violence*, 3, 275-289.
- Gondolf, E.W. (1998, July). *A 30-month follow-up of court-referred batterers in four cities*. Paper presented at Program Evaluation and Family Violence Research: An International Conference, Durham NH.
- Gondolf, E.W. (1999a). A comparison of four batterer intervention systems: Do court referral, program length, and services matter? *Journal of Interpersonal Violence*, 14, 41-61.
- Gondolf, E.W. (1999b). MCMI-III results for batterer program participants in four cities: Less "pathological" than expected. *Journal of Family Violence*, 14, 1-17.
- Gondolf, E.W. & Williams, O.J. (2001). Culturally-focused batterer counseling for African American men. Paper submitted for publication.
- Harrell, A.V. (1991). *Evaluation of court-ordered treatment for domestic violence offenders*. Final report submitted to the State Justice Institute. Washington, DC: The Urban Institute.
- Hastings, J.E., & Hamberger, L.K. (1988). Personality characteristics of spouse abusers: A controlled comparison. *Violence and Victims* 3, 31-48.
- Healey, Kerry, Smith, Christine & O'Sullivan, Chris. (1998). *Batterer Intervention: Program Approaches and Criminal Justice Strategies*. Washington, DC: US Department of Justice.
- Leonard, K.E. & Jacob, T. (1988) Alcohol, alcoholism, and family violence, in VanHasselt, Morrison, Bellack, & Hersen (Eds.) *Handbook of family violence*, 383-406. NY: Plenum.
- Murphy, C.M., Meyer, S.L., & O'Leary, K.D. (1993). Family of origin violence and MCMI-II psychopathology among partner assaultive men. *Violence & Victims*, 8, 165-176.
- Murphy, C.M., Musser, P.H., Maton, K.I. (1998). Coordinated community intervention for domestic abusers: Intervention system involvement and criminal recidivism. *Journal of Family Violence*, 13, 263-284.
- O'Leary, K.D. (1993). Through a psychological lens: Personality traits, personality disorders, and levels of violence. In R.J. Gelles & D.R. Loseke (Eds.) *Current controversies on family violence* (pp. 7-30). Newbury Park CA: Sage.
- O'Leary, K.D., Heyman, R.H., & Neidig, P.H. (1999). Treatment of wife abuse: A comparison of gender-specific and conjoint approaches. *Behavior Therapy*, 30, 475-505.
- Palmer, S.E., Brown, R.A., & Barrera, M.E. (1992). Group treatment program for abusive husbands: Long term evaluation. *American Journal of Orthopsychiatry*, 62, 276-283.
- Pence, E. & Paymar, M. (1993). *Education groups for men who batter: The Duluth model*. New York: Springer.

- Rosenfeld, B. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review, 12*, 205-226.
- Russell, M.N. & Frohberg, J. (1995). *Confronting abusive beliefs: Group treatment for abusive men*. Newbury Park CA: Sage.
- Saunders, D.G. (1996). Feminist-cognitive-behavioral and process-psychodynamic group treatments for men who batter: Interaction of abuser traits and treatment. *Violence & Victims, 11*, 393-414.
- Sonkin, D. J. & Durphy, M. (1997). *Learning to live without violence: A handbook for men*. Volcano CA: Volcano Press.
- Stordeur, R.A. & Stille, R. (1989). *Ending men's violence against their partners*. Newbury Park CA: Sage.
- Syers, M. & Edleson, J.L. (1992). The combined effects of coordinated criminal justice intervention in woman abuse. *Journal of Interpersonal Violence, 7*, 490-502.
- Taylor, B.G., Davis, R.C., & Maxwell, C.D. (2001). The effects of a group batterer treatment program in Brooklyn. *Justice Quarterly, 18*, 170-201.
- Toby, J. (1957). Social disorganization and stakes in conformity: Complimentary factors in the predatory behavior of hoodlums. *Journal of Criminal Law Criminology, and Political Science, 48*, 12-17.
- Tolman, R.M. & Bennett, L.W. (1990). A review of quantitative research on men who batter. *Journal of Interpersonal Violence, 5*, 87-118.
- Tolman, R.M. & Edleson, J.L. (1995). Intervention for men who batter: A review of research. In S. Stith & M.A. Straus (Eds.) *Understanding partner violence: Prevalence, causes, consequences, and solutions* (pp. 262-274). Minneapolis, MN: National Council on Family Relations.
- Weisz, A.N., Tolman, R.M., & Saunders, D.G. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence, 15*, 75-90.
- Williams, O.J. (1995). Treatment for African-American Men Who Batter. *CURA Reporter, 25*, 6-10.
- Williams, O.J., & Becker, L.R. (1994). Partner abuse programs and cultural competence: The results of a national study. *Violence and Victims, 9*, 287-295.

APPENDIX:

BRIEF DESCRIPTION OF SELECT EXPERIMENTAL AND QUASI-EXPERIMENTAL BIP EVALUATIONS

Quasi-Experimental Evaluations

City of 200,000. Chen, Bersani, and Denton (1989) compared 120 convicted batterers in an 8-session BIP with 101 convicted batterers who did not get the BIP. Participants in BIP were half as likely to re-offend during the year after treatment, but only if they attended at least 6 of the 8 sessions.

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

Baltimore County. Harrell (1991), supported by the State Justice Institute, studied batterers ordered to one of three 12-session BIPs, and batterers not ordered to a BIP. The BIPs all differed in their orientation to services. At follow-up, men in the BIP condition (n=81) were more likely to have been physically violent and more likely to have been re-arrested than men in the control condition (n=112).

Pittsburgh, Houston, Dallas, & Denver. Gondolf (1999a, 2000), supported by the Center for Disease Control and Prevention, followed 840 men in four different BIPs, using both court records and partner interviews, for as long as three years after intake. Despite differences in referral (diversion v. conviction), length of treatment (3 to 9 months), or additional services (e.g. mental health, substance abuse) there were no significant differences between the four groups at 15-month follow-up in re-assault (32%), controlling behavior (45%), re-arrest (26%), or victim perception of safety (72%).

Experimental Evaluations

U.S. Navy in San Diego. Dunford (2000), supported by the National Institute of Mental Health, compared outcomes for men randomly assigned to either (a) a 1-year cognitive-behavioral BIP, (b) a 1-year couples group, (c) a rigorous monitoring program similar to assertive probation work, or (d) a safety planning condition approximating a control group. Men with substance abuse problems or mental disorders were excluded from the study. At one-year follow-up, 48 men in the BIP condition had slightly lower incidence of recidivism by partner report than 50 men in the control group, but there were no differences in rate of re-arrest.

Minneapolis. Edleson and Syers (1990) randomly assigned 283 batterers to one of three programs (self-help v. educational v. combined) and one of two program intensities (weekly for 3 months v. twice weekly for 4 months). A six-month follow-up with 92 program completers and their partners found no significant differences between models or intensities, although there was a non-significant trend favoring the educational approaches over the self-help approach.

Broward County, Florida. Feder and Forde (2000), supported by the National Institute of Justice, studied 404 men randomly assigned to either probation plus a Duluth-based BIP or probation only. At follow-up, there were no significant differences between the BIP and the probation-only group in attitudes toward wife beating, attitudes toward women, or self-reported likelihood they would hit their partner in the future. The men in the BIP were less likely to view their partner as responsible for the violence, and as more dangerous to his partner than men in the control group. At one-year follow-up, men in the BIPs were no less likely than men in the control group to be re-arrested for domestic violence.

Hamilton, Ontario. Palmer, Brown, & Barrera (1992), with support from the Ontario Ministry of Community and Social Services, conducted the first experimental evaluation of BIPs when they studied 59 men randomly assigned either to a 10-week psycho-educational group or to no intervention beyond supervision. Recidivism, as measured by official records, was significantly greater for men in the control group.

Madison, Wisconsin. Saunders (1996), supported by the Centers for Disease Control, randomly assigned 218 batterers to cognitive-behavioral or process-psychodynamic group treatments. In 18 to 54-month follow-up with program completers, there were no differences in arrests or in victim-reported violence or fear of violence between the two treatment approaches. However, men who had a higher levels of dependency did better in the psychodynamic treatment, and men who had a more antisocial orientation did better in the cognitive-behavioral program.

Brooklyn. Taylor, Davis, and Maxwell (2001), supported by the National Institute of Justice, compared men randomly assigned to either a 40-hour Duluth-based BIP (n=186) or 40 hours of community service (n=190). Evaluations occurred at 6 months and again at 12 months, and included both official records of complaints/arrests and victim interviews, although only half of the victims could be interviewed. Results suggest significantly lower recidivism for men in BIPs using official reports, but no difference between BIPs and community service using victim report.



In Brief:
**Controversies and Recent Studies
of Batterer Intervention Program Effectiveness**

Batterer intervention programs (BIPs) are designed for men arrested for domestic violence, or for men who would be arrested if their actions were public. The goals of BIPs include not only preventing further violence, but also serve as an adjunct form of justice. BIPs are one local node of a community violence prevention effort. Knowledge about batterer program effectiveness is important because:

- Courts are referring men convicted of domestic abuse to batterer intervention programs, suggesting a certain level of public confidence in the effectiveness of these programs;
- Referral of a batterer to a BIP is one of the strongest predictors that a woman will leave shelter and return to the batterer. Advocates are justifiably concerned that batterer programs not hold out a promise of hope which may become a vehicle for injury;
- People who work with batterers are interested in outcomes so they can improve the level of effectiveness; and,
- Evaluation is one of the best forms of program accountability. A program which evaluates its effect on batterers (and victims) is likely a safer program than one which does not evaluate.

Outcomes for batterer intervention program are still relatively few, but a picture of their effectiveness is beginning to emerge. At the present time, we believe studies suggest the following:

- BIPs have a small but significant effect. Batterers who do not complete their program are twice as likely to be re-arrested. The effect of any of the elements of a community violence prevention effort are equally small, including education, arrest, prosecution, probation, victim services, and adjunct services. The most effective reduction in partner violence will occur in those communities with the strongest combination of coordinated, accountable elements;
- BIPs are more effective for some men than others. One in four men referred to a BIP will account for most of the repeat violence and most of the serious injury within a batterer program. The batterer program alone will not effectively reduce his potential for violence, so the batterer program's best role for these hard-to-treat men may be to hold them in program as long as possible, increasing the time a battered women may need to get herself into a safer position. Although longer term programs may not be more effective in rehabilitating these batterers, they may serve a more useful function as agents of accountability and victim safety than shorter term programs;
- Most re-offense occurs early, usually within six months of initial program intake. Assessment and accountability must be on-going, not something which is done only at program intake and follow-up; and,



- No program approaches have shown themselves to be superior to other approaches, so standards that specify specific BIP models must do so with criteria other than recidivism in mind. Due to concerns about victim safety, batterer programs must not only hold their participants accountable, but BIPs must also hold themselves accountable as well. The best way to do this is to work closely with local victim services agency or victim advocates.

Evaluation of batterer programs is fairly new, and evaluation of the community responses within which batterer programs operate is even newer. Statements about batterer program outcomes can be understood only in the context of their limitations. Specifically:

- While nearly 50 empirical studies have been published on batterer program outcomes, in only four of these studies were batterers randomly assigned to a BIP or a no-treatment control group. Experimental research is difficult and expensive, and is, at present, inconclusive;
- The rate of attrition from batterers groups has been on the order of 50%, and the people most likely to drop out are also most likely to re-offend;
- It is difficult to distinguish the effects of batterer programs from the community of sanctions and victim services in which the BIP operates. Nor should we necessarily try to do so, since such a distinction contradicts the theoretical premises of most batterer programs; and,
- The primary outcome for BIPs-re-offense-is measured in two ways: victim report and re-arrest. Victim contact is always difficult and potentially unsafe. A substantial proportion of batterers will either be living alone or moved on to their next partner. Re-arrest data grossly under-estimates re-offense. Batterer self-report is not considered a valid indicator of outcome.