A comparative multisite evaluation was conducted in four geographically distributed cities to examine the relative effectiveness of different approaches. The intervention systems represented a range of court-referral procedures (pretrial or postconviction), program duration (3 months to 9 months), and additional services (occasional referral or in-house alcohol treatment). 210 men at each site were recruited and tested (background, Millon Clinical Multiaxial Inventory, Michigan Alcoholism Screening Test). The batterers' partners were interviewed by phone every 3 months over a 15-month follow-up after intake, with a response rate of 77% overall. There was no significant difference in the reassault rate, portion of men making threats, and victim quality of life across the four sites. The longest, most comprehensive program did, however, have a significantly lower rate of severe reassault substantiated in a logistic regression controlling for background variables. The findings suggest that differing intervention systems that conform to fundamental standards can achieve similar outcomes.

A Comparison of Four Batterer Intervention Systems

Do Court Referral, Program Length, and Services Matter?

EDWARD W. GONDOLF
Mid-Atlantic Addiction Training Institute
Indiana University of Pennsylvania

The proliferation of batterer programs in the wake of court-mandated counseling has given rise to a variety of intervention systems. Batterer programs vary substantially in their counseling approach, program duration, and extent of services. They also differ in their linkages to the court, victim services, and other community agencies. The few evaluations comparing and evaluating program features pose a mixed array of results. The published evaluations, moreover, do not address a combination of features and tend to focus on

---

Author's Note: The program directors at the four research sites offered essential assistance in developing and implementing the data collection: Robert Foster, Toby Myers, Sherry Lundberg, and Robert Gallup. Several research assistants, especially Jennifer Daly, Jody Scruggs, Don Bozich, and Paul Burchfield, helped to administer questionnaires and test materials at the sites, along with assistance from staff at the respective sites. Jeff Coben, Ron Laporte, and Ed Ricci of the University of Pittsburgh have served as coinvestigators and scientific advisors. Jewel Lee Doherty, project coordinator, Crystal Deemer, administrative assistant, Neil Fulton, data manager, and several data-entry workers, graduate assistants, and student staff of the Mid-Atlantic Addiction Training Institute provided a variety of support, supervision, and service. The Texas

JOURNAL OF INTERPERSONAL VIOLENCE, Vol. 14 No. 1, January 1999 41-61

counseling approach and format. They are also compromised by small samples, unique locales, and either low response rates or varying follow-up periods.

A controlled experiment comparing program format (counseling approach and intensity of sessions) was conducted approximately a decade ago. Edleson and Syers (1990) compared the outcome of a mutual help approach, a didactic approach, and a combined didactic and discussion group in Minneapolis. Each option also included an alternative of 12 sessions over a 12-week duration versus a more intensive 32 sessions over a 16-week duration. Approximately 80% (n = 153) of those initially assigned to a group (n = 283) completed a group. More than half (52%) of these completers (n = 92) were interviewed 6 months after the program. According to victims (n = 80) and batterers (n = 12), the reassault rate was 32% and 38% for the didactic and combined programs, respectively, and 54% for the self-help group. The differences in the rates were not, however, statistically significant. There was also no significant difference between the 12-week and 16-week alternatives within each program option.

More recently, Saunders (1996) compared the outcomes of men randomly assigned to a 16-week didactic group (cognitive-behavioral modality) versus a discussion group (psychodynamic modality) in Madison, Wisconsin. Nearly two thirds (62%) of the 218 assigned men completed their assigned program, and more than three fourths (79%) of the 136 program completers were interviewed between 6 and 54 months after the program. The reassault rate was approximately 50% for both options, according to women's reports, which were adjusted for reassault reported by the women's batterers and the batterers' police records. (The reassault rate based on the unadjusted reports of only the initial female partners was 30%.) Additional experimental evaluations are currently in progress examining a particular program approach against a control group (see Davis & Taylor, in press, for a review).

Several single-site evaluations have examined short-term programs and long-term programs, programs with different counseling approaches, and different court-referral procedures (Tolman & Bennett, 1990). However, it is difficult to compare these studies and the different components they examine because the samples, instruments, and methodologies differ. Furthermore,

---

Council on Family Violence, the Pennsylvania Coalition Against Domestic Violence, the Mid-Atlantic Addiction Training Institute, and the Center for Injury Research and Control at the University of Pittsburgh Medical Center all contributed to the development of the research proposal and to the coordination of its implementation. The research was made possible through a grant from the Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (Grant No. R49/CCR310525-02), but does not necessarily represent the official view of the CDC.

the experimental studies, although useful in identifying the effects of modality or format, are confined to single sites that may not be representative of other communities, may experience sample bias as a result of noncompliant subjects in random assignment to treatment options, and may employ idealized options that are not readily adopted by practitioners. Lastly, the previous evaluations tend to neglect the broader system of court procedures, batterer counseling, and additional services that comprise batterer intervention. This tendency may inadvertently attribute outcomes or effects to a particular batterer program approach that may, in fact, be related to the intervention system as a whole.

We conducted a naturalistic, comparative evaluation of multiple sites in an effort to address some of these conceptual and methodological limitations and, in the process, to further the research on the relative effectiveness of different batterer intervention systems. The design is naturalistic in the sense that it examines programs as they are implemented in their respective communities, and it is comparative in the sense that systems of a range of components and linkages are selected. Our main objective was to more specifically examine the impact of differences in court referral, program duration, and extent of services on reassault rates. We did, however, consider a number of outcome measures for what amounted to a 15-month follow-up from program intake.

We initially hypothesized that more comprehensive systems with postconviction, longer-duration, and extra services would have significantly lower reassault rates than less comprehensive systems. Batterer program participants are frequently characterized as having social reinforcement that needs to be offset by decisive sanctions and observation and as having individual problems (e.g., alcohol abuse and personality disorders) that warrant additional specialized treatment (Hart, Dutton, & Newlove, 1993; Tolman & Bennett, 1990). The more comprehensive systems include added sanctions imposed by postconviction as opposed to pretrial referral, extended batterer counseling and observation implied in longer-term programs, and the additional alcohol treatment and victims services that address compounding problems.

METHOD

Research Sites

Four batterer intervention systems were selected to approximate a continuum of court referral, program duration, and additional services while

TABLE 1: Components of Batterer Intervention Systems

	Intervention System			
Component	*Pittsburgh*	*Dallas*	*Houston*	*Denver*
Court referral	Pretrial with court review	Postconviction	Postconviction	Postconviction
Program duration	3 months	3 months	6 months	9 months
Additional services	Referral	Assessment and referral Women's and children's groups	Assessment and referral Women's contact Women's groups	Evaluation and in-house substance abuse treatment Individual mental health counseling Women's casework

conforming to prevailing program standards in the field. The first condition of selection was that the systems be based on a well-established batterer program. By well-established, we mean that the programs (a) comply with their respective state standards for batterer programs, (b) collaborate with battered women's services in their respective communities, (c) employ a curriculum that approximates the cognitive-behavioral approach presented in the prevailing published manuals in the field (e.g., Kivel, 1992; Pence & Paymar, 1993; Russell, 1995; Stordeur & Stille, 1989), (d) have been in operation for 5 years or more with at least 40 to 50 referrals per month, and (e) conduct training and supervision of new and branch programs. These programs contrast with programs that may be newly formed or in transition, experimental or atypical in their approaches, or not actively involved in linkages with other services. (Cognitive-behavioral approaches focus on restructuring thought patterns, including rationalization, minimization, justification, and on teaching self-talk and other techniques to interrupt specific behaviors.)

Second, intervention systems were selected that represent a range of components (see Table 1). At one extreme is what could be considered the least comprehensive intervention. The Pittsburgh system relies on pretrial referrals, requires 3 months of weekly sessions, and makes referral for court-identified substance abuse or mental problems. The system in Denver represents the most comprehensive combination of components. It relies on mandatory sentencing to counseling as part of conviction, 9 months of weekly group counseling, and includes an extensive clinical evaluation, in-house alcohol treatment sessions, individual psychotherapy for mental problems,

TABLE 2: Additional Services Delivered at Research Sites

	Site						
Services	*Pittsburgh* % (n)	*Dallas* % (n)	*Houston* % (n)	*Denver* % (n)	*Total* % (n)	χ^2 (df)	*Significance*
Men							
Any alcohol treatment	17 (33)	21 (36)	23 (40)	29 (52)	22* (161)	8.88 (3)	.031*
Other counseling	16 (31)	19 (28)	20 (31)	28 (47)	21 (137)	7.79 (3)	.051*
Women							
Any shelter contact	29 (53)	37 (52)	38 (60)	48 (80)	38 (245)	12.78 (3)	.005**
Other counseling	26 (47)	31 (44)	35 (54)	53 (89)	36 (234)	30.69 (3)	.001***
Any alcohol treatment	18 (33)	20 (28)	13 (20)	23 (38)	18 (119)	5.90 (3)	.117

NOTE: Pittsburgh = 3-month, pretrial program; Dallas = 3-month, referral program; Houston = 6-month, contact program; Denver = 9-month, in-house program.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

and women's case management. The Dallas system used postconviction referrals to 3 months of discussion-oriented sessions, with individual assessments, individual counseling, and women's groups available in addition to the batterer group sessions. The Houston system was based on a postconviction, 6-month, didactic program with efforts to contact and offer support groups to the battered women, as well as to make referrals for apparent batterer substance abuse problems. The pretrial procedures in Pittsburgh include a court liaison from the batterer program and a 30-day court review in a designated domestic violence court. The court liaison briefly orients men referred to the program and sets a date for program intake; the court reviews confirm the men's compliance to and completion of the program (see Gondolf, 1998).

The delivery of additional services during the program appears to confirm the continuum (see Table 2). The men at the Denver site were approximately 10% more likely to receive alcohol treatment or other counseling (beyond the batterers counseling) than the Pittsburgh men, and the Denver women were nearly twice as likely to receive battered women services or additional counseling than the Pittsburgh women (see Table 2). Alcohol treatment includes any detoxification, inpatient, or outpatient treatment, and any participation in

Alcoholics Anonymous (AA) meetings. Other counseling includes individual couples, group (not AA or Narcotics Anonymous), clergy, or mental health counseling; and battered women's services refers to any overnight stay in a women's shelter, attendance at battered women's counseling or support groups, or phone advice on a hotline or with battered women's program staff. (The portion of men in need of alcohol or psychological treatment was similar across the sites, according to an alcohol-screening test and a psychological inventory discussed below.)

The continuum of interventions may, however, be confounded by variations in components. Our periodic observations of program groups and interviews with program staff revealed much more complexity in program procedures, structure, organization, and staffing than originally anticipated (for further details, see Gondolf, 1998b). Programs varied in the time between court date and program intake, staff supervision and coordination, and staff training and facilitator styles. For instance, the average time from court to intake was 2.5 weeks for the pretrial system in Pittsburgh and 3 months for the postconviction system in Houston. The communities also differed in police response to domestic violence, media attention to domestic violence, and social service availability. For instance, the rearrest rates are substantially lower in Dallas even though the percent of women calling the police during follow-up is similar across the sites (Gondolf, 1997).

Subjects

The first 20 to 25 men appearing for program intake each month were recruited as subjects until approximately 210 men had been recruited at each of the four research sites (total sample $N = 840$). The refusal rate was less than 5% at each of the four research sites, suggesting a representative sampling of the program referrals for the recruitment year of 1995. At program intake, trained research assistants administered a background questionnaire and forms soliciting written consent to participate in the program evaluation. Additionally, the Millon Clinical Multiaxial Inventory, Version III (MCMI) (Millon, 1994) was administered to assess psychopathology among the subjects (base rate [BR] score ≥ 75), and the Michigan Alcoholism Screening Test (MAST) (Selzer, 1971) was used to assess alcohol abuse (MAST total score ≥ 5).

The vast majority of the men (82%) were mandated to the programs by the courts, as opposed to others who entered the programs voluntarily (18%). The demographics of our sample of batterers reflect the prevailing generalizations of batterers in court-mandated batterer programs. The men tended to

be in their early thirties (M = 32 years; SD = 8.8) and of lower socioeconomic status (64% blue collar). There were, however, contrasting portions of men within demographic categories. About half (55%) of the men were from racial minorities versus half (45%) from the White majority, a fourth (24%) of the men were without a high school education versus a third (36%) of the men who had some college education, two thirds (64%) of the men were fully employed versus a third (36%) who were underemployed (part time or unemployed), and half (49%) of the men were living with their partners and the other half (51%) were not (see Gondolf, 1996, for further details and documentation on the sample characteristics). Overall, the batterers in our sample are more racially diverse, less likely to be married, and more likely to be unemployed than batterers described in studies conducted in the Midwest (Hamberger & Hastings, 1988) and Canada (Dutton, 1986).

The subsamples for each site had similar rates of men who had previously abused or injured their partners, had been previously arrested, had parents who abused them or who had drinking problems, had alcoholic tendencies (according to the MAST), and had displayed symptoms of severe psychopathology (according to the MCMI) (Gondolf, 1998b). The variables for relationship status (i.e., marriage, children, living together) were also equivalent for the four subsamples. The site subsamples did significantly differ in terms of demographics (e.g., Pittsburgh had the highest portion of men with lower socioeconomic status and Denver the lowest) and initial referral sources (i.e., 94% of the Pittsburgh men were court referred, whereas 77-79% of men at the other sites were court referred). The postconviction systems, as opposed to the pretrial system of Pittsburgh, had longer time lapses between the court hearing and program intake and consequently had a larger portion of men with new partners and men who had no contact with their initial partners.

Follow-Up

The program outcome is based primarily on phone interviews with the initial or new partners of the men in our sample. Research assistants attempted to call these women within 2 weeks of their partners' program intake to verify the batterers' background information, and then every 3 months over a 15-month follow-up period. Letters were sent requesting women without phones or who could not be reached to call a toll-free number, and the respondents were paid $10 to $20 for each completed interview. A female partner was interviewed for 77% of the batterers at program intake and at least once during the 15-month follow-up. The response rate varied, however, across the

four sites between 70% and 85% (85% at Pittsburgh, 71% at Dallas, 75% at Houston, and 80% at Colorado).

A female partner was interviewed for 65% of the batterers over at least 9 months of the follow-up period (interviewed for three successive 3-month follow-up periods). Of the initial victims, 72% were contacted for at least a portion of the follow-up period, and 57% were interviewed for at least 9 months. A new female partner was interviewed for 13% ($n = 107$) of the batterers; both a new partner and the initial victim were interviewed for 8% ($n = 68$) of the men in our sample. The new partners were identified through follow-up interviews that were conducted with the men to gain information to address other research questions.

The women respondents were, for the most part, representative of the base sample of cases ($N = 840$). There was no significant difference between the batterers' characteristics of the interviewed women and of those not interviewed in terms of demographics, relationship status, alcohol abuse, mental health problems, and criminality. However, minority women were less likely to be in the follow-up sample (49% of the respondents vs. 63% of the nonrespondents were minority women, $\chi^2[1] = 15.09$, $p \leq .001$). This difference may, in part, be because African American women were less likely to be married to their partners and less likely to be living with them during the follow-up. Also, the women in the follow-up were more likely to be partners of men who completed the program (66% of the respondents vs. 55% of the nonrespondents completed at least 3 months of the batterer programs; $\chi^2[1] = 9.49$, $p \leq .002$). This completion difference, in part, reflects the fact that the African American men were more likely to drop out of the programs (47% of the African American men, 42% of the Latino men, and 31% of the Anglo men dropped out; $\chi^2[3] = 17.70$, $p \leq .001$).

This racial and dropout bias does not appear to have affected the research results according to two additional analyses. First, the African American women who remained in the follow-up were reassaulted at approximately the same rate as White women (33% vs. 35%). Second, a capture-recapture analysis estimated a reassault rate for the entire sample, including missing cases, comparable to the reassault rate derived from only the responding cases (Gondolf, Chang, & LaPorte, in press).

Outcome Variables

The primary outcome used to compare the intervention systems is the reassault reported by the women partners (initial and new partners) of the

subjects from each site during the 15-month follow-up. Reassaults were assessed through a series of questions that included an open-ended question about how the relationship was going, descriptions of any conflicts and their circumstances, an inventory using the categories of the Conflict Tactics Scale (CTS) (Straus, 1979), the nature of battering injuries and medical assistance received for those injuries, and the woman's response to the abuse. A reassault was considered any incident that included one of the tactics on the physical aggression subscale of the CTS (i.e., push, shove, grab; slap; hit with a fist, bite, kick; hit with something, attempt to hit with something; choke or burn; threaten with a knife or gun; use a knife or gun; force sex against will). Severe reassault is defined as use of any of the so-called severe tactics of the CTS (i.e., hit with a fist, bite, kick; hit with something, attempt to hit with something; choke or burn; threaten with a knife or gun; use a knife or gun; force sex against will). The women's reports were used as the basis for reassault rates, because batterers have been shown to substantially underreport their assaults and, consequently, most previous program evaluations have relied on the women's reports (Edleson & Brygger, 1986). The reliability of the women's reports in our sample is substantiated in a separate document (Heckert & Gondolf, 1997).

Several other forms of abuse are used as outcome measures as well. Inventories for controlling behaviors (i.e., kept from talking on the phone, kept from spending time with friends, stopped from going someplace, followed against will or knowledge, kept from using income or savings, taking money from the person), verbal abuse (i.e., swore or screamed at; accused of other relationship; put down, insulted, called names), and threats (i.e., threatened to hit, attack, or harm the woman; kill any person; take away or harm children; harm other people; kill or hurt self) were converted into variables that indicated a "yes" response to any of the possible items in that category of abuse during any of the follow-up interviews. These inventories were compiled from several abuse scales that attempt to identify nonphysical abuse (e.g., Marshall, 1992; Shepard & Campbell, 1992; Straus, 1979; Tolman, 1989).

The women's subjective appraisal of their own well-being was also considered. The women were asked a global question: "Would you say that your life is generally better, worse, or the same?" The women were additionally asked to estimate how safe they felt at that point and how likely it was that their partners would hit them in the next few months (using a Likert-type scale). The responses from the last follow-up interview (at the end of the 15-month follow-up) were used for these variables.

Analysis

To test for the site effect on outcome, outcome variables were cross-tabulated by site (i.e., Pittsburgh, Dallas, Houston, Denver) using the chi-square-test statistic. The site variable was also collapsed to compare the longest, most comprehensive program (Denver) against the shortest, least comprehensive program (Pittsburgh). Two-site comparisons were tabulated to explore for significant differences that might have been obscured by the four-category site variable. As previously mentioned, the samples were conveniently similar in terms of previous abuse, prior arrests, parental abuse, alcoholic tendencies, psychopathology, and relationship status. These similarities provide some control of influential factors that might account for differences in outcome and make the cross-tabulations at least instructive.

To test whether certain batterers might have better outcomes in one site or another, three-way cross-tabulations for reassault were also computed controlling for variable categories of batterers. Three categories of clinical concern and interest were identified: (a) court referrals only (vs. noncourt referrals), (b) 3 months of program attendance (vs. program dropouts by 3 months), and (c) repeat reassaulters (vs. men who committed one reassault during the follow-up). The court-referral category is derived from the self-report of the subjects at intake about the primary referral source. The 3-months attendance category is used to accommodate the minimum program duration of 3 months, and because 3 months appears as a threshold for dropout in the longer programs—that is, the dropout rates do not substantially increase ($\leq 10\%$) in the longer programs after 3 months. (The 3-month dropout rate was 31% for Pittsburgh, 40% for Dallas, 45% for Houston, and 36% for Denver.) The repeater category refers to men who physically abused their partners more than once during the 15-month follow-up, according to the partners' reports of reassault. Cross-tabulations were also computed, controlling for reassault reported by just the initial woman partner and for women who still had contact with their partners.

Two logistic regressions were computed to test for a site effect on any reassault and for severe reassault while controlling for a combination of background variables and the influential variable for referral source identified in the cross-tabulations. The variable for being frequently drunk (i.e., once a month or more), according to the women's report, was used as a measure of alcohol abuse instead of the MAST or the alcohol dependency subscale of the MCMI, because the Pittsburgh site had a significantly higher percentage of men identified as drunk (73% in Pittsburgh vs. 46% in Houston; $\chi^2[3] = 27.29$, $p \leq .001$), and the MAST results did not significantly differ

across the sites. When the MAST is entered in into the equation, it has no significant effect.

The MCMI results indicating narcissistic and antisocial tendencies (BR ≥ 75 on these respective subscales) and severe pathology (BR ≥ 75 on any "severe personality pathology" or "severe syndrome" subscale) were also entered as behavior characteristics, because they have been associated with reassault in previous studies (Hamberger & Hastings, 1990). These subscales were replaced with the results of other MCMI subscales and also with four batterer types based on a factor analysis of the MCMI (Gondolf, 1998c), with no effect on the outcome of the equation.

We first tested for significant covariance among the background variables through cross-tabulations and did not find any substantial covariance among the designated variables (Gondolf, 1996). The background variables were then entered into the equations as a block, followed by dummy variables for the research sites. The Pittsburgh site, which was least comprehensive in duration and services, was used as the base comparison group and was therefore not entered as a dummy variable. An additional set of regressions was computed, contrasting the longest, most comprehensive Denver site with the other three sites combined. These regressions explore the effect of the longest, most comprehensive program relative to the other programs and test the influence that sample size may have on producing significant results. (We also compared results for other combinations of sites with less distinctive results—that is, the two longer programs against the two shorter programs, and the longest against the two shortest programs.)

The logistic regressions are based on 568 cases, which result from cases deleted because of missing data in the 658 cases of women respondents. The missing data are the result of using women's information about the men's alcohol abuse and previous severe assault as background variables. Many of the women did not know about the men's behavior in this regard, or some were not available at intake when the background information was collected. The women's reports about the men are preferred because of the men's substantial underreporting on these variables. The equations were recomputed using the men's reports as proxies for missing data in these two variables. This procedure enabled the inclusion of the full sample of women respondents and produced similar results to those reported here.

FINDINGS

The rates for reassault and the other 15-month outcomes are relatively equivalent across the four sites, with one exception. There is tentative evi-

dence of a lower rate of severe reassault at the 9-month program. The rate for severe reassault is significantly lower for the longest, most comprehensive program in the bivariate analysis for the research sites. This potential site effect approaches significance when controlling for difference in background variables through a logistic regression. Other indicators for a site effect (e.g., the increase in the goodness-of-fit and correct-classification rate) are, however, very weak.

Table 3 compares the outcomes for the four sites in terms of reassault, other abuse, legal action, and the women's well-being. The reassault rate (i.e., any reassault) for the longest program is 8% to 9% less than for the two 3-month programs, but this difference is not statistically significant. The severe reassault and repeated assault rates at the two 3-month programs are, however, twice that of the 9-month program—this difference is statistically significant. Only 12% of the women responding for the site with the longest and most comprehensive program (Denver) reported severe reassault, whereas 23% to 26% of the women from the two shortest programs reported severe reassault, ($\chi^2[3] = 11.70$, $p = .008$). This difference in severity is not directly reflected in the portion of women reporting injuries. The portion of women injured was 13% for the women with partners in the 9-month program, 19% at the pretrial, 3-month program, and 26% at the postconviction, 3-month program. The difference between the 9-month and 3-month postconviction program is statistically significant, ($\chi^2[1] = 7.05$, $p = .008$), but not the difference between the pretrial, 3-month program.

Any differences among the sites in terms of other abuse and legal action are minimal. The variables for women's well-being present a contradictory outcome picture. As expected, women associated with the longest batterer program are slightly more likely to feel they will not be hit again (77% vs. 63% for other sites; $\chi^2[1] = 6.45$, $p = .011$), but there is no significant difference in the percentage who report feeling very safe. Also, the women in the shorter programs were more likely to report being better off than those in the longest program (67% to 70% vs. 59%; $\chi^2[1] = 5.17$, $p = .023$). This outcome is the opposite of what was expected.

Table 4 shows reassault rates by site while controlling for different categories of batterers and victims. These results confirm, for the most part, the trends suggested in the initial bivariate cross-tabulations of Table 3, but also indicate that differences in the men's previous behavior and contact with their female partners may contribute to the reassault rates among the sites. Reassault rates converge for those men not previously arrested for a crime other than domestic violence and for those not drunk at least once a month. In other words, the reassault rates for men without compounding problems are even more similar across the four sites than the rates for the full respondent

TABLE 3: Comparison of Outcome for Four Batterer Intervention Systems

	Research Sites											
Outcome	*Pittsburgh*		*Dallas*		*Houston*		*Denver*		*Total*			
Variable	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	χ^2 (df)	*Significance*
Reassault												
Any reassault	35	(63)	36	(52)	30	(48)	27	(47)	32	(210)	3.93 (3)	.270
Severe reassault	23	(41)	26	(38)	21	(33)	12	(21)	20	(133)	11.70 (3)	.008**
Repeated reassault	24	(43)	24	(34)	16	(25)	11	(19)	19	(121)	13.03 (3)	.005**
Injury	19	(35)	26	(37)	21	(33)	13	(23)	19	(128)	8.03 (3)	.046*
Other abuse												
Controlling behavior	43	(78)	47	(67)	47	(75)	46	(79)	45	(299)	.69 (3)	.874
Verbal abuse	77	(140)	69	(101)	64	(102)	68	(117)	70	(460)	7.69 (3)	.053
Threats	45	(81)	42	(60)	45	(72)	40	(70)	43	(283)	1.01 (3)	.783
Legal action												
Police called	21	(38)	22	(31)	24	(38)	23	(39)	23	(146)	.65 (3)	.885
Any rearrest	26	(45)	28	(39)	28	(42)	21	(33)	26	(159)	2.82 (3)	.421
Women's well-being												
"Very unlikely" to be hit	66	(84)	60	(52)	62	(62)	77	(83)	66	(281)	8.19 (3)	.042*
Feel "very safe"	77	(105)	70	(62)	71	(79)	72	(83)	72	(343)	1.73 (3)	.633
Quality of life												
Better	67	(92)	68	(60)	70	(78)	59	(68)	12	(54)		
Same	25	(34)	15	(13)	23	(26)	24	(27)	22	(100)		
Worse	8	(11)	17	(15)	7	(8)	17	(20)	12	(54)	12.90 (6)	.045*

NOTE: Pittsburgh = 3-month, pretrial program; Dallas = 3-month, referral program; Houston = 6-month, contact program; Denver = 9-month, in-house program.
*$p \leq .05$. **$p \leq .01$.

sample. The rates also converge when considering only women who still have personal contact with the men at the end of the follow-up period.

The reassault rates for men who have attended at least 3 months of program sessions (i.e., 3-month completers) are, however, much lower in the longest, most comprehensive program (22% vs. 35%; $\chi^2[1] = 5.39$, $p = .020$). This difference is substantial and statistically significant, but may be influenced by the required court review at the 3-month, pretrial program. The court review may keep high-risk men from dropping out at the 3-month

TABLE 4: Reassault Rates Controlling for Different Categories of Batterers and Victims

	Research Sites						
Category	*Pittsburgh* % (n)	*Dallas* % (n)	*Houston* % (n)	*Denver* % (n)	*Total* % (n)	χ^2 (df)	*Significance*
Male batterers							
Court-referred only	35 (63)	28 (33)	27 (36)	25 (40)	29 (172)	4.33 (3)	.228
3-month completers	35 (45)	32 (28)	21 (19)	22 (25)	28 (117)	8.15 (3)	.043*
Not previously arrested	28 (28)	37 (29)	24 (21)	24 (21)	28 (99)	4.63 (3)	.201
Not drunk (monthly)	23 (10)	26 (9)	30 (18)	20 (12)	25 (49)	1.43 (3)	.697
Female victims							
Initial victim only	33 (59)	37 (47)	28 (41)	26 (44)	31 (191)	4.53 (3)	.210
Personal contact	37 (50)	42 (38)	31 (29)	26 (28)	34 (145)	6.09 (3)	.107

NOTE: Pittsburgh = 3-month, pretrial program; Dallas = 3-month, referral program; Houston = 6-month, contact program; Denver = 9-month, in-house program.
*$p \leq .05$.

program. The reassault rates do not substantially change when considering only the court-referred men (i.e., excluding the voluntary program participants) or the initial victims only (i.e., excluding the outcomes of any new partners).

Table 5 shows the results of logistic regressions for any reassault and severe reassault during the follow-up period. The type of intervention system is not a substantial or significant predictor of any reassault when demographics, relationship status, behavioral characteristics, and referral source are controlled. In the equation for severe reassault, the coefficient ($\beta^2 = -.635$) and odds ratio (Exp[*B*] = .530) for the longest, most comprehensive program (Denver) suggest an association in the expected direction and approach statistical significance ($p = .065$).

The site variables do not significantly improve the goodness-of-fit for any reassault ($\chi^2[3] = 2.88$, $p = .411$) or for severe reassault ($\chi^2[3] = 6.80$, $p = .078$). Neither equation has much predictive power, even with all the background and site variables included. The equation for any reassault correctly classified only 31% of the actual reassault cases, and the severe reassault equation classified only 9% of the actual severe reassault cases. In sum, the logistic

TABLE 5: Logistic Regression of any Reassault and Severe Reassault (*N* = 568)

	Any Reassault			*Severe Reassault*		
Variable	*Beta*	*Exp* (B)	*Significance*	*Beta*	*Exp* (B)	*Significance*
Demographics						
Age			.077			.154
18-25	.581	1.788	.040*	.608	1.837	.065
26-35	.089	1.093	.711	.191	1.211	.506
36-65	—	—	—	—	—	—
Race			.129			.194
Anglo	—	—	—	—	—	—
African American	–.189	.828	.417	–.291	.748	.275
Latino	–.697	.498	.031*	–.784	.457	.037*
Other	–.643	.526	.184	-.364	.695	.511
Education			.455			.068
11th grade or lower	.265	1.303	.349	.733	2.080	.020*
High school graduate	.270	1.310	.237	.325	1.385	.233
Some college	—	—	—	—	—	—
Employment			.910			.620
Employed full-time	—	—	—	—	—	—
Employed part-time	–.086	.917	.744	–.064	.938	.834
Unemployed	–.102	.903	.718	.269	1.308	.390
Relationship status						
Married	–.063	.939	.769	.085	1.089	.729
Living with partner	.306	1.358	.130	.392	1.480	.093
Children at home	–.548	.578	.006**	–.487	.615	.035*
Personality traits (MCMI)						
Narcissistic tendencies	–.332	.718	.167	–.065	.937	.810
Antisocial tendencies	.184	1.202	.462	.168	1.183	.554
Severe pathology	.370	1.448	.110	.399	1.490	.122
Behavioral characteristics						
Drunkenness (monthly)	–.090	.914	.719	.392	.676	.162
Previous severe woman assault	.844	2.326	.001***	1.044	2.840	.001***
Previous arrests (not DV)	.461	1.586	.022*	.611	1.842	.008**
Other referral source (not court)	933	2.541	.004**	.409	1.505	.265
Intervention system (Site)			.419			.094
3-month, pretrial (Pittsburgh)	—	—	—	—	—	—
3-month, referral (Dallas)	.005	1.005	.987	.216	1.241	.521
6-month, contact (Houston)	–.094	.910	.758	–.016	.984	.963
9-month, in-house (Denver)	–.414	.661	.148	–.635	.530	.065
Constant	–2.553	—	.001***	–2.845	—	.001***

NOTE: MCMI = Millon Clinical Multiaxial Inventory.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

regressions show no substantial site effect on any assault and, at the very most, a very weak site effect on severe reassault.

A site effect for the longest, most comprehensive program became more evident when the site variable contrasted the longest, most comprehensive program against the other three programs combined. The coefficients and odds ratios for all the background variables and the longest program remained virtually the same, but the program coefficient ($\beta^2 = -.386$) and ratio (Exp[*B*] = .680) further approached statistical significance ($p = .102$) in the regression for any reassault and became significant ($p = .017$) in the regression for severe reassault ($\beta^2 = .659$, Exp[*B*] = .499). The change in the significance levels is a response to the increased size of the site comparison group made possible by collapsing the three shorter programs together. In other words, we might be more confident with a larger sample that the results are not merely the result of chance.

In response to the revised site variable, the goodness-of-fit also improved slightly for the any reassault ($\chi^2[1] = 3.03$, $p = .082$) and severe reassault ($\chi^2[1] = 6.14$, $p = .013$) equations. The correct classification decreases to 25% for any reassault and to 8% for severe reassault.

CONCLUSION

Interpretations

We initially hypothesized that intervention systems based on longer programs and more comprehensive services would have lower reassault rates. The predominant finding in our comparison of batterer intervention systems is the relatively weak evidence for a substantial difference across the four sites in our study. The 15-month reassault rates are relatively similar for the four sites, despite differences in batterer demographics, court linkages, and program format. The finding regarding any reassault persists when controlling for postprogram variables (e.g., the man's drunkenness after the program, contact with the initial partner, and additional assistance and intervention) in a longitudinal analysis of the follow-up data (Jones & Gondolf, 1997). The outcomes in terms of other forms of abuse, legal action, and women's well-being are also nearly equivalent overall.

There is, however, a significant trend of lower rates of severe and repeated reassault in favor of the most comprehensive program. This potential site effect approaches significance in the logistic regression controlling for background variables and becomes significant when the longer program is

contrasted to the other three programs combined. This effect may be related to the alcohol treatment available in the longest, most comprehensive program, because alcohol abuse is often associated with more severe assault (Kantor, 1993).

There are three main qualifications that may account for these unexpected findings. First, the slight trend favoring the longer and more comprehensive programs may become more apparent in an extended follow-up. There may be a threshold beyond our relatively short, 15-month follow-up when reassault begins to reescalate. Second, more systematic implementation and specific evaluation of alcohol treatment may further differentiate the programs and their outcomes. This is of special concern given the apparent influence alcohol abuse during follow-up has on reassault (Jones & Gondolf, 1997) and that only 12% more of the men received alcohol treatment at the longest, most comprehensive program as compared to the shorter, least comprehensive program (29% vs. 17%; see Table 2). Third, the expected outcomes may not be apparent because of other system components that neutralize the benefits of longer and more comprehensive programs. For instance, the quick response and mandatory court-review employed in the 3-month, pretrial program may offset some of the consequences of its shorter program duration and less extensive services (see Gondolf, 1998). Fourth, some components of an intervention system simply may not be implemented as effectively or fully as they could be. For example, probation offices may be slow or inconsistent in their response to noncompliance, and women's services may not reach the victims most at risk for reassault.

An additional issue with outcome evaluations of this sort is the clinical significance, as opposed to statistical significance, of the findings (Morrison & Henkel, 1970). The differences in reassault rates (in the cross-tabulations) or the site effects on reassault (in the logistic regressions) may matter in the "real world," even though they are not statistically significant in our analyses. For instance, the fact that 8% more men reassaulted their partners at one site as opposed to another is of major consequence to the victims at that site. Even though the difference is not statistically significant, it may be clinically significant. Moreover, the site samples are arguably nonrandom, because they are selected at a unique time from unique programs. Probability statistics, on which significance levels are based, do not necessarily apply in this case. The magnitude of the rate differences and site effects is what is important. The longer programs are associated with at least some rate difference and site effect. Our impression is that they are not substantial, given the extent of variation across the intervention systems.

Programmatic Implications

Our finding of nearly equivalent reassault rates, on the surface, seems to reinforce the arguments for brief therapy and managed care (Budman & Gurman, 1988). The short-term, pretrial, didactic program appears to be the most efficient system. The numbers of men enrolled in the pretrial program were two to four times that of the other programs, and with a budget substantially less per man than the other 3-month, discussion-oriented program. The relative effectiveness of the 3-month, pretrial program may also raise some questions about the recent push for more extensive individual assessments and specialized treatment for men who batter. This program did only preliminary screening for apparent drug and alcohol problems and mental health problems in a group-orientation session and watched for behavioral problems in the subsequent group sessions. Other programs in the study had individual sessions to assess, evaluate, or test the men for alcohol and mental health problems.

Our findings do not, however, necessarily indicate that one intervention system is better than the others. The intervention systems may simply have similar essentials. The systems in this study, as discussed in the methods section, were all based on well-established batterer programs. They complied to state standards for batterer programs, which require that batterers be held accountable for their abuse, that rationalizations for abuse be exposed, that woman battering be identified as a means of power and control, and that woman battering not be attributed to stress or substance abuse. They also collaborated with battered women's services and their respective criminal justice systems.

It may be that each intervention system is defined more by the composite of its components and experience than by its individual components. Rather than as part of a continuum of systems, each system may be a unique adaptation to a peculiar set of resources, leadership and staffing, court procedures, and community expectations. Each appears to have organically developed with its own history and internal culture. Therefore, one system or another may not be readily replicated in, or transferred to, another community that has a different set of constraints and opportunities. In sum, some of these systems may not be appropriate or suitable for another community and must be viewed as peculiar to their own locale.

Summary

Court linkage, program duration, and extent of services may not, in themselves, influence program outcome. These components may interact with, or

even be superseded by, other components, such as delays from court hearing to program intake or inconsistent probation response to noncompliant program participants. The comparable results of the pretrial, shorter term, didactic program raise at least the possibility of expanding options within the court. The findings lend tacit support to accepting referrals through limited screening instead of extensive clinical evaluation and to accepting men who may not be visibly motivated at intake. This option appears to be relatively effective when embedded in a larger system of supportive women's services and reliable court responses to noncompliance.

The influence of either external community factors or essential program components could, of course, be tested in a multisite experimental design, comparing different program formats at each site. One format may produce lower reassault rates regardless of site, or different formats may perform better at different sites, as is the case in multisite experimental studies of depression treatments and alcohol treatments (Elkin, Shea, & Watkins, 1989; Mattson & Allen, 1991). Unfortunately, the complexity, ethical problems, and cost of such a design may be prohibitive. Furthermore, the push toward coordinated community responses to battering makes batterer programs increasingly a part of some larger intervention system. To account for this, program evaluation may need to become more akin to systems analysis. Our comparison of different intervention systems hopefully offers a step in this direction.

REFERENCES

Budman, S., & Gurman, A. (1988). *Theory and practice of brief therapy*. New York: Guilford.

Davis, R., & Taylor, B. (in press). Does batterer treatment reduce violence? A synthesis of the literature. *Women and Crime*.

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

Edleson, J., & Brygger, M. (1986). Gender differences in self-reporting of battering incidents. *Family Relations, 35*, 377-382.

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

Elkin, I., Shea, M. T., & Watkins, J. T. (1989). The National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*, 971-982.

Gondolf, E. (1996, November). *Characteristics of court-mandated batterers in four cities*. Paper presented at the annual meeting of the American Society of Criminology, Chicago, IL.

Gondolf, E. (1997). Patterns of reassault in batterer programs. *Violence and Victims, 12*, 373-387.

Gondolf, E. (1998a, February). *The impact of mandatory court review on batterer program compliance: An evaluation of the Pittsburgh municipal courts and Domestic Abuse Counseling Center (DACC)*. Report submitted to the Pennsylvania Commission on Crime and Delinquency, Harrisburg, PA.

Gondolf, E. (1998b, June/July). Do batterer programs work? A 15-month follow-up of a multisite evaluation. *Domestic Violence Report*, *3*, 64-65, 78-79.

Gondolf, E. (1998c). Batterer types based on the MCMI: A less than promising picture. Unpublished paper, Mid-Atlantic Addiction Training Institute, Indiana, PA.

Gondolf, E., Chang, Y., & LaPorte, R. (in press). Capture-recapture analysis of batterer reassaults: An epidemiological innovation for batterer program evaluation. *Violence and Victims*.

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

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

Hart, S., Dutton, D., & Newlove. T. (1993). The prevalence of personality disorders among wife assaulters. *Journal of Personality Disorders*, *7*, 329-341.

Heckert, A., & Gondolf, E. (1997, June/July). *Agreement of men's and women's reports of woman assault during batterer program follow-up*. Paper presented at the Fifth International Family Violence Research Conference, Durham, VT.

Jones, A. S., & Gondolf, E. (1997, June/July). *Post-program predictors of reassault for batterer program participants*. Paper presented at the Fifth International Family Violence Research Conference, Durham, VT.

Kantor, G. K. (1993). Refining the brushstrokes in portraits on alcohol and wife assaults. In S. Martin (Ed.), *Alcohol and interpersonal violence: Fostering multidisciplinary perspectives* (NIAAA Monograph No. 24) (pp. 281-290). Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Kivel, P. (1992). *Men's work: How to stop the violence that tears our lives apart*. Center City, MN: Hazelden.

Marshall, L. (1992). Development of the severity of violence against women scales. *Journal of Family Violence*, *7*, 103-121.

Mattson, M., & Allen, J. (1991). Research on matching alcoholic patients to treatments: Findings, issues, and implications. *Journal of Addictive Diseases*, *11*(2), 33-49.

Millon, T. (1994). *Millon Clinical Multiaxial Inventory: III*. Minneapolis, MN: National Computer Systems.

Morrison, D., & Henkel, R. (1970). *The significance test controversy*. Chicago: Aldine.

Pence, E., & Paymar, M. (1993). Education groups for men who batter: The Duluth model. New York: Springer.

Russell, M. N. (1995). *Confronting abusive beliefs: Group treatment for abusive men*. Thousand Oaks, CA: Sage.

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.

Selzer, M. (1971). The Michigan Alcoholism Screening Test (MAST): The quest for a new diagnostic instrument. *American Journal of Psychiatry*, *3*, 176-181.

Shepard. M., & Campbell, J. (1992). The Abusive Behavior Inventory: A measure of psychological and physical abuse. *Journal of Interpersonal Violence*, *7*, 291-305.

Stordeur, R., & Stille, R. (1989). *Ending men's violence against their partners*. Newbury Park, CA: Sage.

Straus, M. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) scales. *Journal of Marriage and the Family*, *41*, 75-88.

Tolman, R. (1989). The development of a measure of psychological maltreatment of women by their male partners. *Violence and Victims*, *4*, 159-177.

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

Edward W. Gondolf, Ed.D., M.P.H., is associate director of research at the Mid-Atlantic Addiction Training Institute (MAATI) and professor of sociology at Indiana University of Pennsylvania.