Evaluating Interventions With Violent Offenders: A Guide for Practitioners and Policymakers

BY PATRICIA VAN VOORHIS, FRANCIS T. CULLEN, AND BRANDON APPLEGATE

Growing frustration with the Nation's high rate of violence has prompted a sustained movement to "get tough" with violent offenders. Policies ranging from mandatory minimum sentences to "three strikes and you're out" promise to increase both the number of dangerous offenders who will be incarcerated and the duration of their time behind bars. In reality, however, many violent offenders will receive community-based sanctions, particularly during the early years of their criminal careers. And even if imprisoned, many will return to society.

Indeed, increased use of incapacitative and deterrent approaches is at best a partial solution to violent crime. Although prison typically is warranted for other reasons, the factors associated with violent behavior are not likely to be affected by after-the-fact prison terms (Gendreau & Little, 1993; Quinsey & Walker, 1994). Clearly, correctional agencies continue to be challenged to implement treatment programs that can be targeted to violence.

Unfortunately, a "knowledge gap" exists on the details of prevention and treatment programs that might be most effective with violent offenders. To date, important steps have been made in understanding the principles of effective correctional treatment programs; we also have learned a good deal about what types of programs "work" (Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990; Gendreau & Ross, 1987; Lipsey, 1988; Palmer, 1992). Yet, existing "meta-analyses" of evaluations of treatment programs, which identify program factors that are important in reducing recidivism, do not cite many studies specifically targeted to violent offenders (see Andrews et al., 1990b; Gendreau & Ross, 1987; Lipsey, 1988; Palmer, 1992; Whitehead & Lab, 1989).

It is noteworthy that a recent National Academy of Sciences Panel on the Understanding and Control of Violence confirms the continuing knowledge gap in the treatment of violent offenders and underscores the need for more effective evaluations of existing programs (Roth & Raiss, 1993). The panel has issued important conclusions:

a. Findings of program evaluations are not yet conclusive enough to warrant a national commitment to any single strategy.

b. Strategies aimed at predisposing risk factors, even when they are effective, require time to demonstrate that they will work.

c. While some strategies will doubtless prove more effective than others, the diversity of violent events guarantees that no single strategy will prevent more than a small fraction of them.

d. Violence control policy should be committed to small investments in the testing of many small-scale but sustained problem-solving initiatives—each initiative focused on a specific source of violence.

The panel has recommended, further, that each treatment initiative should involve five steps:

a. Diagnose the problem, using criminological and epidemiological techniques to document its importance and identify risk factors that suggest a preventive strategy.

b. Develop prototypes of several tactics for strategy implementation that show promise based on theory, research findings, or experience.

c. Compare the effectiveness of the alternative tactics through rigorous evaluations that use randomized assignment wherever feasible.

d. Refine the tactic for implementation using the evaluation findings as the basis.

e. Replicate the evaluation and refinement steps to sharpen the effectiveness of the interventions and adapt them to local community characteristics.

Past research confirms the panel's wisdom. Most of what we have learned has come from small, well-controlled studies, rather than from large-scale initia-
tives. We often neglect to replicate the most successful programs, and we often fail to create a clear prototype that will facilitate replication. The paucity of evaluation research, and the failure to replicate successful strategies, embodies the tragedy of many social programming endeavors: Most of what we know to be effective is not currently in practice; many of the seemingly good ideas which constitute current practice have not been or are not being tested.

Closing the knowledge gap on what works with violent offenders, thus, will depend on the development and effective evaluation of programs within a variety of settings. Yet, conducting effective evaluations often is viewed as a time-consuming and frustrating challenge to administrators and practitioners.

We approach this challenge from the growing realization that most of the impediments to conducting a useful evaluation occur far in advance of the program evaluation during the design phase of a program. And many of these design problems then adversely affect the evaluation and its findings. Design flaws may even affect whether the program can be evaluated.

The intent of this article is to discuss the interrelationship between program design and program evaluation. We review a number of program issues that become impediments to conducting sound evaluations of correctional interventions. We then offer numerous suggestions. Ultimately, we hope to put forward reasonable programmatic and evaluation directions which will then increase the likelihood of finding positive evaluation results. We do not address those activities that the evaluators are either solely responsible for or can execute with minimal input from the staff. These include development of a research design and comparison groups, data analysis, and report preparation. The issues put forward in the sections below, however, are of more primary importance because technical expertise is of little value otherwise.

**Impediments to Conducting Successful Program Evaluations**

Unfortunately, in the field of corrections, it is the rare evaluation that is not doomed far before the first data collection form is completed. As the following sections detail, evaluation impediments begin in the environment of the program being tested and emerge from the manner in which the program is designed. At key points, administrators and program staff may be setting themselves up for unnecessary program disappointments.

**Problems Within the Program Climate**

Perhaps one of the first questions we might ask is: Is it safe for staff and administration to evaluate their program? One does not have to look far to observe that evaluation results have been misused to undermine the continued existence of correctional interventions. Evaluation results are extremely vulnerable to political interests and are often misinterpreted. They are too frequently conducted with the understanding that negative evaluation results will sound the final bell for a program. Indeed, at one point in our recent history, misinterpreted evaluations placed the entire goal of rehabilitation in jeopardy and resulted in severe fiscal cuts to state and Federal treatment budgets (e.g., see accounts of the Martinson (1974) "nothing works" debacle as detailed in Cullen & Gilbert, 1982). Small wonder that program staff may not want to cooperate with evaluation research. A more constructive approach is sorely needed.

Evaluation efforts are also sometimes marred by a lack of financial and organizational support. This occurs in many ways. Agency administrators may not be willing to endorse or might withdraw support from the research design, compromising the integrity of the comparison group or the random assignment procedures in the middle of the research. Staff may not be able to break free from ongoing responsibilities to meet in planning or evaluation sessions. Finally, there may be insufficient financial resources to conduct the research. In such situations, programs may want rapidly assembled results that can be shown to their stakeholders (e.g., referral sources, funders, and political constituencies). And while they may be able to produce an evaluation quickly, the result may be an evaluation that: a) is based upon remote measures of program success, b) cannot be used to learn much of anything about effective programming, and c) is even more likely, by virtue of the program and evaluation design flaws, to portray the program as a failure.

Perhaps the most important issue concerns how well the program has been planned by the administrators and staff and how well program staff are then able to articulate various program components to evaluators. As a recent panel on Violence Prevention for Young Adolescents observed, evaluations must be planned at the earliest stages of program design (Wilson-Brewer, Cohen, O'Donnell, & Goodman, 1991). Here, the mindset of a good evaluation and the mindset of good program planning are strikingly similar. The program that cannot describe the population it serves, or is not delivering a clear process of treatment, or is not targeting the treatment to a cause of the problem being treated, usually will not succeed. Beyond that, it usually should not be evaluated because evaluation results will be highly misleading.

In response, the "evaluability assessment" has become an increasingly popular practice among government agencies contemplating funding a substantial evaluation of an experimental initiative (Rossi & Free-
man, 1989). An evaluability assessment involves an external consultant’s examination of a program’s plan. A decision against evaluating the program is often made in instances where program staff and planners cannot identify several key questions: Who are our clients and what problem characteristics are we treating? What intervention fits our client problems? Why did we choose this particular intervention? What are our goals and objectives? What intervention fits these goals and objectives? How will we know when we have implemented this intervention according to design? How will we know whether this intervention was effective?

Often, corrections is pressured to use a less rigorous planning process. Daily, the theoretical “magic bullets” are put forward by the media and by politicians; such pressures drive policy and intervention designs (Palmer, 1992). The proposed panaceas often prescribes an intervention that: a) does not target a cause of crime; b) has already been found to be ineffective (e.g., boot camps without a treatment component); c) has no theoretical reason for working (e.g., yelling at boot camp participants); or d) denies everything that we know about crime and its effective treatment. As a result, most of what we know about effective interventions is not currently in practice; an emerging knowledge base that shows great consistency across authors and studies is virtually being ignored (see Andrews et al., 1996b; Andrews & Bonta, 1994; Gendreau & Ross, 1987; Lipsey, 1988; Palmer, 1992; Van Voorhis, 1987). The following sections expand upon these issues.

Who Are We Treating?

The identification of key client characteristics is important for two reasons. First, in order to successfully treat a client for violence or aggression, we must point (or target) our program services to some individual traits/problems or environmental factors that are known to predispose individuals to violence. It is not overly difficult to identify the risk factors for aggression and violence, for we probably know far more about the causes of aggression and violence than we know about their treatment. Aggressive or antisocial behavior is typically the result of an interaction between personal characteristics and situational factors (Goldstein & Keller, 1987; Quinsey & Walker, 1994). In a valuable summary of the causes of violence, the National Advisory Panel accounts for both macro social (e.g., poverty, physical structure of a neighborhood) and micro social (e.g., family and social disorganization, bystander or activity) correlates of violence. These are portrayed as interacting with numerous psychosocial and biological factors (Roth & Reiss, 1998). Equally instructive is a causal sequence of events shown in the work of Arnold Goldstein and his associates at Syracuse University (see Goldstein & Keller, 1987), where the authors identify perceptual and cognitive patterns, coping skills, contingencies, and values as ideal targets for intervention.

The second set of individual characteristics speaks to clients’ amenability to treatment. Even when targeting our services to those clients at risk of violence/aggression, additional factors such as motivation, personality, and intelligence will affect a client’s success in the program, regardless of whether or not they are correlates of violence.

As will be seen shortly, knowledge of the second set of individual characteristics (treatment amenability factors) affects both the plans for service delivery and the evaluation. Program staff will need to facilitate offenders who are likely to have difficulties in the program, while evaluators should want to consider these factors in their analyses of the evaluation results. Andrews and Bonta (1994) include these characteristics under their “responsivity principle,” while Warren (1983) and Palmer (1978, 1984, 1992) incorporated them earlier under the notion of differential treatment or “matching.” An important program that is doing very little diagnostic assessment work or is not screening for specific risk factors is likely to be making poor decisions about what specific aspect of violent behavior might be addressed by the program. Programmatic decisions about risk and responsibility factors should be related to the services the program delivers, its criteria for program admission, and the assessments it conducts at admission.

When we do not accommodate program interventions to offender risk and amenability factors, the following are the most likely outcomes:

a. A good program fails because it is targeted to individuals who cannot benefit from the program.

b. A good program works with some and not with others. Our successes are canceled out by our failures. The program looks bad when it really did work with some offenders.

c. The program was a true failure because planners did not provide a service which targeted a factor that is related to violence.

Did We Choose and Implement an Effective Intervention?

When evaluators, program personnel, planners, and others are not able to articulate the type of treatments or services delivered by a program, we refer to the programmatic services as a “black box.” We do not, in other words, know what is in the box. In fact, it is not uncommon to observe evaluations with outcome data (e.g., recidivism measures or improvements on test
scores, but no clear indication of what the program did to achieve these results. In such situations we know what the program accomplished, but we do not know what it did.

The seriousness of this problem is underscored by a view to the earlier history of evaluation research and correctional treatment. As the term suggests, some of our earlier published evaluations showed outcomes, but they failed to note whether the program was able to do what it was designed to do. Moreover, some programs provided results, but perhaps never had a clear intervention model. Thus, it is entirely possible that many of the failures of correctional treatment were not failures of a program but rather programs that never occurred.

There are several ways in which a program “might not occur”:

- The program is not grounded in the knowledge base of the discipline and therefore is not utilizing program strategies that are either empirically or theoretically sound.
- The program has chosen a strong program design but is not operating according to the design or the clinical dimensions of an intervention (e.g., a social learning program which does not incorporate principles of good role modeling).
- There is a lack of specificity in the program’s design; we are delivering some global treatment, e.g., counseling, case management, job skills (which can mean different things to different staff), rather than an explicit treatment process that is known to be effective.
- Staff members do not understand the intervention. They do not wish to cooperate, or they do not have or do not follow a treatment manual.
- The organizational and political climate is too confusing and is not conducive to successful implementation.
- Budget cuts create a situation where we are asked to do the impossible—keep the program without the funds.
- The “dosage” is inadequate. It may be a good intervention, but the amount of time the client participates in the program is insufficient.
- Clients did not or could not attend.
- The program is too turbulent, undergoing several changes during the evaluation. Evaluators do not know what they tested.

The most tragic outcome of such events occurs when an evaluation creates the impression that “nothing worked” when, in fact, “nothing happened.” The importance of adequate program implementation was recently illustrated poignantly by the Violent Juvenile Offender Project. The project had been implemented in four sites but only showed success in two. The factor that differentiated success from failure was the quality of program implementation, which the evaluators measured meticulously (Fagan, 1990).

There are other problems with neglectful interventions. They create situations where we cannot make program adjustments or replications because we don’t know what the program did.

Will We Know If We Have Succeeded?

The answer to this question involves constructing measures of program success. Very often “success” is measured by using general measures of recidivism, such as revocation, rearrest, or reconviction. However, while recidivism, especially for new violent offenses, is an important measure of program success, programs should not ignore a second type of measure linked to intermediate objectives. These depict whether or not our program affected a risk factor for aggression/violence. In answering the earlier question, “who are we treating?,” we might have indicated that we were treating some “cause” of aggression (e.g., poor conflict management skills). In this case, an intermediate objective would address improvement in conflict management skills. Usually we would measure attainment of this objective just before each client’s release from the program.

Unfortunately, the importance of intermediate objectives and measures of their attainment are often overlooked, particularly in programs for violent offenders (Wilson-Brewer, Cohen, O’Donnell, & Goodman, 1991). In doing so, two problems are created. First, the program ignores an important measure of treatment integrity. And stakeholders are entitled to ask, “Why isn’t the program addressing aggressive behavior?” In contrast, an obvious way to show that the program had targeted aggression and had delivered an effective intervention would be to show that it affected risk factors related to aggression.

Second, the program ignores another function of these measures; if properly chosen, they tap dynamic risk factors. Dynamic risk measures offer extremely important information to programs because the research shows us that improvement on dynamic risk factors is typically as predictive or more predictive of
post program success than are more traditional risk assessment instruments (Andrews & Bonta, 1994).

The task of obtaining the more distal outcome or “success” measures of recidivism is more complicated, particularly with respect to violence or aggression (see Gullone & Keller, 1983; Monahan & Steadman, 1994). And each type of measure has its own source of measurement error. Medical and psychological interventions, for example, often use analogue measures which pose problems of external validity (i.e., a subject's aggressive response to a hypothetical situation may not translate into actual aggression). Self-report measures may be marred by behavioral and temporal specificity, social desirability, response bias, attribution styles, and motivation. There is also a tendency to pick up minor forms of aggression, rather than more serious behaviors. Staff observational measures, which are very efficiently obtained in institutional settings, may reflect subject reactivity to the observer or staff biases. Measures of overt violent behavior (official measures) may result in high amounts of undetected behavior and low base rates (Monahan & Steadman, 1994; Van Voorhis, 1994a). Finally, each measure is likely to tap a different component of aggression, ranging from overt aggression to aggressive attitudes to aggressive tendencies.

In contrast to the measures of intermediate objectives and recidivism, program completion rates, client satisfaction surveys, or number of clients served are not measures of program success. Although they serve some purpose in the larger scheme of program accountability, these measures do not answer any questions about “what works” in the treatment of the violent offender.

Suggestions for Improving Program Evaluation and Planning

Improving the Organizational Climate

Any improvements in the organizational support for evaluation research must affect positive change on two fronts: the organization's perspective on the evaluation and the staff's skill in planning and developing programs. With respect to the former, some improvements have already occurred, largely in response to increased pressures on agencies for accountability from funding sources. Indeed, there are good reasons for conducting an evaluation. We might, for example, the American Correctional Association's accreditation criteria which, while not requiring research, nevertheless offer additional credit for participation in evaluations or other types of research. In another sense, an evaluation, if properly conducted, can help a program improve its planning and effectiveness. In doing so, an evaluation can engage staff in a constructive share of the planning. Staff involvement in planning and program articulation can be a source of staff motivation. In support of this notion, one of the characteristics of an effective program is the absence of top-down planning and the presence of staff involvement in program design (Andrews & Kieckling, 1980; Gullen & Gendreau, 1992; Gendreau, in press).

We can expect substantial improvements in the climate of a program evaluation by simply rejecting the “win-lose” perspective on evaluation research, which sees programs as either “working” or “not working.” If the evaluation study is comprehensive enough, it will produce more information than whether the program worked. It may, for example, identify specific program components or services that failed (rather than the entire program); it may show us that the program worked for some types of offenders but not for others; it may tell us whether the program achieved a proper service “dosage.” If none of these components of the program are measured in the evaluation, these crucial questions will not be answered. When they are assessed, however, the evaluation may become an important source of feedback, thus leading to program adjustments rather than to program obliteration (Rossi & Freeman, 1989).

We might also create a fairer climate for the program by promoting more widespread acceptance for realistic standards of success. Indeed, more conservative researchers tell us that the most we can hope for, with the most effective program designs and implementation, is a 20 percent improvement in recidivism rates for an experimental group over a comparison group (see Lipsey, 1988; Palmer, 1992), perhaps 40 to 50 percent among the most optimistic reviewers (see Andrews & Bonta, 1994; Gendreau, in press). While some have argued that these figures indicate failure rather than success (see Lab & Whitehead, 1990; Martinson, 1974), even the conservative 20 percent success rate translates into impressive cost benefits (Gendreau & Ross, 1987; Lipsey, 1984). More importantly, the 20 percent success figure for treatment is the best we have because alternative policies of incapacitation and deterrence don’t come close to this figure (see Irwin & Austin, 1994; Roth & Reiss, 1993), and often the deterrent or “get tough” strategies find higher recidivism in the experimental programs than in the comparison groups (Gendreau & Ross, 1987).

A Model for Program Planning

Even when organizational considerations are addressed successfully, sound planning must become a structural component of programming at both administrative and staff levels of responsibility. Much of what we have addressed in this article comes together in the diagram shown in figure 1, which also depicts
the relationship between the program design and the evaluation design.

![Diagram](image)

**Figure 1. The Interrelationship Between Planning and Evaluation Tasks.**

The program begins with a thorough understanding of the target problems and the characteristics of individuals who need to be served in order to address these problems. Program selection/screening criteria follow from this understanding. At the same time, the evaluation staff designs and uses intake data collection forms that tap the selection criteria, any relevant assessment data, and routine demographic and social background factors.

Given the staff members' understanding of the problem and their client base, the program identifies objectives that it can reasonably meet within its current resources. There will be intermediate objectives, depicting what the program accomplishes, and long-term objectives, such as a post-program reduction in recidivism. The evaluation staff helps the program to identify measures or data that must be collected in order to assess attainment of these objectives. Often change in one of the intake assessment measures will be one such measure, e.g., improvements on the Buss-Durkee Hostility Inventory.

Finally, when the program designs its service delivery strategy (e.g., a skills development program), the evaluation staff helps the program to devise methods for determining whether the service was administered according to design and how much time clients spent in the program (durance). These factors will be assessed through service delivery measures.

Imposing our knowledge of who and what we are treating

Figure 1 indicates a box for “target problems.” This is where any programmatic intervention begins. Objectives, the choice of interventions, and ultimately measures of success all relate back to the problems we are addressing. If we do not know what the problem is, it is impossible to plan clear objectives, interventions, and outcome measures. Clearly, most programs for aggressive individuals should plan for some level of client assessment.

Goldstein and Keller (1987), for example, provide clear but intensive examples of program planning that moves from a clear assessment of the problem to interventions that fit each problem. The authors assess adult clients on each of the types of target problems that they address and give examples of psychological tests and inventories that fit each target problem: 1) for measuring arousal-heightening interpretations, the Anger Inventory (Novaco, 1975) or the Reaction Inventory (Evans & Strangeland, 1971); 2) for assessing heightened affective arousal, the Buss-Durkee Hostility Index (Buss & Durkee, 1957) or marital conflict measures (Filsinger, 1980); 3) for malcommunication, the Conflict Tactics Scale (Straus, 1979); 4) for deficiencies in social skills, the Social Performance Survey Schedule (Lowe & Causton, 1975) or the Conflict Resolution Inventory (Hartwig et al., 1980); and 5) for assessing antisocial values, the Psychopathy Checklist (Hare et al., 1992) or the Sociomoral Reflections Measure (Gibbs & Widaman, 1982).
EVALUATING INTERVENTIONS

From there, the authors proceed to link specific interventions to each problem area. For example, they recommend that clients known to have distorted perceptions of what is a provocative event receive Anger Control Training (see also Novaco, 1975). Likewise, clients with deficiencies in social skills should engage in skills training, etc.

For those who view Goldstein's approach as an overly elaborate model, a more straightforward method is found in Andrews' notion of the 'criminogenic need' (Andrews, Bonta, & Hoege, 1996); Andrews & Bonta, 1994). Many programs have been introduced to needs assessment instruments that supplement the current risk assessment forms (e.g., see NIC, 1982). However, the standard needs assessment often neglects close attention to the 'criminogenic need,' which involves identifying that risk factor which is linked to a particular offender's criminal actions.

For programs or the aggressive offender, this could involve a checklist of aggressive risk factors incorporated into a routine intake interview and staff discussions of the criminogenic needs at team meetings and staffings.

We might also consider the use of certain assessment instruments to make determinations about program "responsibility" that allow us to identify which type of offender was most successful in the intervention. For these purposes, consider instruments listed by Goldstein and Keller (1987): a personality typology (e.g., the Jezeske Inventory [Jezeske & Wedge, 1983], Quay's behavioral categories [Quay & Parsons, 1972; Quay, 1983; Quay, 1984], or the Megargee MMPI-Based Taxonomy of Adult Offenders); or a theoretically based risk assessment instrument such as the Level of Supervision Inventory (which also addresses criminogenic needs) (Andrews, 1982).

In sum, we mentioned that programs have two purposes in understanding their target population: 1) targeting the services to specific client risk factors and 2) the "responsivity principle," ensuring that even the service that logically fits a client's need also fits his or her abilities, motivation levels, and personality considerations (Andrews et al., 1994; Goldstein & Keller, 1983; Palmer, 1978; Warren et al., 1966). More careful attention to both is likely to greatly improve program and evaluation results.

Improving Treatment Integrity

Problems related to treatment integrity occur in two ways. The first problem occurs when a program is unwilling or unable to operate from the knowledge base of the discipline. Asking the question, "Where did the idea for our intervention come from?" would be extremely important. With growing knowledge about what works, it makes little sense to follow a hunch. Thus, before asking whether a program does work, we must ask whether there is any reason why it should work. Programs grounded in a theoretical model (theory of crime) or found in previous controlled evaluations to be effective really should work, so we at least should start at that point.

Even when adopting a sound program design, it would be easy to fail to communicate a clear picture of the clinical or procedural components to the staff who must administer it. To address this concern, written treatment manuals would appear to be essential documents in efforts to improve the consistency of service delivery models within programs as well as across them (Gendreau, in press). A good treatment manual would provide a "how to" description of the intervention model; it should not be written to serve the dual purpose of a program's promotional brochure.

The fact that programs seldom develop and use treatment manuals may be partly the responsibility of the research and scholarly community who has written the evaluation studies over the years. Seldom are program accounts written in a manner that facilitates replication and the development of a treatment manual. The few exceptions include, but are not limited to, Palmer's and Warren's accounts of the Community Treatment Project (see Palmer, forthcoming; Warren et al., 1966; Warren, 1983); Vicky Agee's Closed Adolescent Treatment Program (Agee, 1979) and the Paint Creek Youth Center (Agee, 1987); Fabiano, Robinson, & Purpuro's (1981) work on Cognitive Skills Training Programs; Don Andrews' accounts of effective social learning approaches (Andrews & Bonta, 1985); and Goldstein's treatments for aggression (Goldstein & Keller, 1987) and Novaco's Anger Control Training (1975) have already been mentioned.

In order to assess the second problem, whether a program is being administered according to design, a fairly straightforward process may be followed. This task involves determining the data elements needed in order to assess program integrity. Some of the more common data needs at this point are:

a. Attendance lists and case management logs to collect valuable information on dosage.

b. Surveys of clients and staff in order to determine the extent to which key program components were adhered to. A skills development program, for example, might survey for whether skills were actually modeled.

c. Intake assessments relevant to the nature of the aggression dynamic (e.g., anger management) could be re-administered immediately following the completion of an intervention. Use of alternative test forms is preferred.

d. Intermediate program objectives should lead to evaluation variables or measures of program integrity. They can only do this if they are writ-
ten in a specific manner. Such clear objectives also reduce program confusion and conflict and facilitate successful program implementation.

Less common measures of program integrity should also be considered. For programs that operate from a more complex clinical model, most forms of family therapy, for example (Van Voorhis, Braswell, & Morrow, 1992), clinical observation or content analysis would be very important. A discussion among family members lacks the clinical dimension of some of the family therapy models known to be effective with some offenders.

Existing assessments of program characteristics are also useful. The Correctional Program Assessment Inventory (CPAI) (Gendreau & Andrews, 1994), for example, offers measure of program integrity and other characteristics. A unique contribution of the CPAI is that it asks program staff members where the idea for their interventions originated (from theory or research). It also makes note of whether a treatment manual is in use.

Improving Measures of Program Effectiveness

A program for violent offenders will ultimately be held to the goal of reducing violent and/or aggressive behavior, but this can be shown by measuring attainment of intermediate objectives (e.g., improvement in conflict management skills) or by measuring future aggressive behaviors or violent crimes (i.e., long-term objectives)—or by both (see figure 1). While the most convincing outcomes pertain to follow-up recidivism for new aggressive offenses, it is surprising, nevertheless, to observe programs neglecting the importance of measuring the attainment of intermediate objectives.

Indeed, full attention to whether intermediate objectives had been achieved is more likely to show a well-designed program in a favorable light than not. This occurs because: 1) programs often are more likely to show success on the intermediate objectives than on long-term, post-program measures because, in most instances, the latter are affected by factors that are outside of the program’s control (Quinsey & Walker, 1994); 2) measurement problems are more likely with the long-term recidivism measures than with measures relevant to intermediate objectives; 3) the attainment of intermediate objectives can be used as both measures of program effectiveness and a measure of treatment integrity; 4) some measures can also be viewed as “dynamic risk factors” which are predictors of recidivism (Andrews & Bonta, 1994); and 5) we may compare recidivism figures for those who succeeded in the program (scored well on the intermediate objective) to those who did not (scored poorly). It is surprising how often evaluation results are presented in an undifferen-
other articulated. To simplify the numerous points made in this article, appendix A restates most in the form of a program checklist or steps that might be taken to ensure a sound program plan and a sound evaluation.

In many respects, our comments need not be limited to the violent offender, but are pertinent to other offender treatment programs or for that matter to evaluations of most social programs. This is true for the structure of our suggestions but not necessarily for their specific content. As for the structure of program planning and evaluation, figure 1, and many of the comments pertinent to it, represent a standard model for the social sciences. But we are less comfortable asserting that specific assessments and strategies for violent offenders can be extended to other offender groups. We are especially not ready to agree with the suggestion that violent offenders and other high risk offenders always can be treated in the same groups (e.g., Genéreau, in press; Quinsey & Walker, 1994), with so few tests of programs specifically targeted to aggression and violence, such assertions appear to be premature. In fact, aggression and violence have been unique risk factors which are not at work in other forms of criminal behavior.

In addition to hampering the development of knowledge, many programs appear to be hurting themselves by not following a sound program and evaluation plan and by conducting evaluations too haphazardly. This occurs in many ways. First, in agreement with Palmer (1975), we suspect that many evaluations showing no treatment effect have, instead, masked the treatment effect. That is, many programs have worked with some offenders and not others, but with no way of subdividing the research sample into subgroups that account for different levels of responsivity, we produce successes which are canceled by our failures. Second, many program evaluations have not fully utilized the results that have maximum chance of showing success (e.g., attainment of intermediate objectives); or they produce evaluation measures which do not fit program services and clients. Third, when program disorganization results in staff members working in different directions or when we are uncertain about the characteristics of program clients, we will probably have insignificant results. Statistical tests are designed to reject all but those events which are most likely to be related, and random events are not related. Moreover, invalid measures attenuate research findings. Fourth, many failures have resulted from treatment factors such as self-esteem, which are not necessarily a cause (or correlate) of aggression. Finally well-supported theories of violence and aggression, and the few extant evaluation studies of effective programs that can point us to programs that work, are not used as guides for program development. We hope we have shown that all of these unfortunate shortcomings can be addressed during the planning phases of the program.

NOTES

1. Comparison groups, particularly those that utilize random assignment of clients to experimental and comparison groups, are essential if the evaluation is to contribute to the knowledge base. Once external evaluation services have been contracted for, however, it is the evaluator's responsibility to design the research itself-staff may be asked to see only that the design is adhered to and may be given instructions as to how to adhere to a specific assignment process.

2. A crucial concern here is for whether the choice of interventions reflects prevailing knowledge about effective programming, or was the intervention selected for some other reason.

3. How, for example, do results differ for the two groups of individuals, those amenable to the program and those who were not?

4. An adequate program evaluation would compare the recidivism measures of an experimental group to a randomly assigned or matched comparison group of participants who did not receive the program services being tested. Given the difficulty of conducting such a "controlled" study, many evaluations compare post program measures to pre-program measures or provide only a post program figure. The latter tells us very little about whether or not the program succeeded. Indeed, it is widely accepted that only controlled studies will produce a contribution to our understanding of the treatment of aggression.

5. These do not exhaust all of the assessments suggested by Goldstein and Kafer (1957). The book also contains a list of assessments that would be suitable for children. Moreover, the authors identify a number of suggestions for designing interviews and surveys pertinent to these problem dimensions.

6. The change in the assessment measure (from intake to program completion) measures attainment of an Intermediate objective, but it could be used as a dynamic risk factor (Andrews & Bonta, 1984) for its impact on the long-term objective of reducing recidivism.

7. This is because evaluators have more control over the creation of measures for intermediate objectives and their administration. In contrast, official measures of new offenses (e.g., police data) are likely to have problems with both validity and low base rates (Monahan, 1983). Moreover, researchers are experiencing increasing difficulties securing access to the national data bases for criminal records.

REFERENCES


Agee, V. (1981). The treatment program: Paint Creek Youth Center. Unpublished manuscript obtained from Paint Creek Youth Center, Bainbridge, OH.


