The Evidence for Violence Prevention Across the Lifespan and Around the World: Workshop Summary

Leigh Carroll, Megan M. Perez, and Rachel M. Taylor, Rapporteurs; Forum on Global Violence Prevention; Board on Global Health; Institute of Medicine; National Research Council

🛒 Add book to cart	🔍 Find similar titles

➕ Share this PDF f 🐦 su in

Visit the National Academies Press online and register for...

✓ Instant access to free PDF downloads of titles from the

- NATIONAL ACADEMY OF SCIENCES
- NATIONAL ACADEMY OF ENGINEERING
- INSTITUTE OF MEDICINE
- NATIONAL RESEARCH COUNCIL

✓ 10% off print titles

✓ Custom notification of new releases in your field of interest

✓ Special offers and discounts

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

by those same state government agencies now, and our unit received such a request, I could respond by going to the Campbell Collaboration website to find a vetted systematic review relevant to the issue of violence prevention. Or I could go to the websites for one of the evidence-based registries and identify some programs that these external groups have vetted as an effective violence prevention strategy, and offer several of these to the Governor, the AG, or some other esteemed requester. Now this could be done in a matter of a few hours.

II.6

MAKING AND USING LISTS OF EMPIRICALLY TESTED PROGRAMS: VALUE FOR VIOLENCE INTERVENTIONS FOR PROGRESS AND IMPACT

Patrick H. Tolan, Ph.D.[9]
University of Virginia

The development of standards that can reliably guide interventions and policies for affecting violence can be one of the most critical steps in reducing the rate of the many forms of violence (HHS, 2001). Systematic and soundly rendered identification of a roster of programs (or sets of practices) that can be relied upon for violence reduction can help streamline efforts while increasing benefits (Sherman et al., 1997; Elliott and Tolan, 1999). Another consideration is that relying on scientific evidence for programming and policy is now generally valued. A reference to evidence-based or empirically tested work now has currency. However, because there is still a considerable lack of consensus about what these terms mean, there is pliability in what they represent. There is increasing reference to "evidence based," but growing uncertainty about what that term means.

To create reliable standards, there must be a scientifically sound and objective determination of what can be considered evidence based/empirically tested. With this accomplished, the field could be given clear understanding of which programs are efficacious (able to be effective), which programs are known to be ineffective (soundly evaluated with no significant benefits or negative effects), and which programs lack determination (mixed results from sound evaluations or no sound evaluations). Thus, such a resource can enable funders, implementers, and policy makers to readily access programs that are most likely to be beneficial (Sherman et al., 1997). This resource would be useful without requiring consumers to have extensive knowledge

---

[9] The author appreciates provision of source material from Delbert Elliott and Sharon Mihalic. However, the views presented are those of the author and are not official representations of the Blueprints Initiative.

of evaluation methods or the specifics of each evaluation. In addition to improving consumer capability, this approach brings violence prevention evaluation in line with scientific standards used in other areas of public health and social welfare to determine efficacy. As the standard becomes more widely used and respected, it can also help inform program developers and funders about the design characteristics needed to validly test effects of programs. In turn this will expedite development of new programs and the breadth of approaches that can be used to reduce violence. Additionally, reliable lists can enable efficient use of funding because development and implementation requirements would be known, saving funds and time when compared to untested and unspecified programs or local initiatives developed de novo.

This paper outlines the rationale and important criteria for developing a practical, efficacious list for violence reduction and prevention, and notes critical challenges in developing a useful approach. It focuses on Blueprints for Healthy Youth Development (http://www.colorado.edu/cspv/blueprints), formerly the Blueprints for Youth Violence Prevention, which uses scientific standards that permit reliable understanding of program effects, connects standards to rationale for level of endorsement, and has been constructed with assertive surveying of potential programs and consistent application of stated standards. It was one of the first such efforts to develop scientifically based standards and program listing. Thus it exemplifies what listings can offer the field. In addition to a descriptive summary, ongoing practical and methodological issues and limitations of the current policies and practices of the Blueprints are discussed. This examination of Blueprints as an exemplar initiative is provided to highlight the advantages such an approach can offer the violence prevention field and to argue for extending this approach and the standards used in Blueprints to other areas of global violence prevention.

The Value of List Making

The need for scientifically sound guidance for violence prevention has long been recognized (Krug et al., 2002). Many scientific studies can be relevant to identify programs that can work, ranging from case studies and qualitative investigations, to trend analyses and representative surveys, to comparisons of groups and conditions for variation in the extent of problems. However, these are all correlational studies, which are informative but cannot provide information on the causal impact of a particular program or set of practices on violence (Tolan and Guerra, 1994). Such research may point to important targets or suggest processes for program emphasis. However, intervention evaluation requires a method that scientifically and quantitatively compares (1) the effects of the program

and (2) not being exposed to the program, with confidence that these differences are only due to different exposures/conditions that potentially allow exposure. Three promising approaches are available for determining what can reduce or prevent violence: systematic reviews across studies (meta-analysis); experimental methods (RCTs); and carefully designed, implemented, and controlled quasi-experiment (not randomly assigned) methods (Shadish et al., 2002).

Although considerable debate exists about developing lists of programs versus systematically identifying key features or practices through meta-analyses, there are several reasons lists may provide the best guidance (Advisory Board of Blueprints for Violence Prevention, 2011; Valentine et al., 2011). One advantage is that, even among the most tested programs, there are few reports, so creating rosters of programs rests on relatively few studies per program. In contrast, other approaches, such as identifying a practice common to effective programs or measuring average effects across multiple studies of similar programs, are susceptible to unstable estimations and reduction of important differences into large categories that cannot direct practice (e.g., cognitive behavior approaches). Despite the large burden borne by violence, funding for violence research is severely limited, particularly for trials of different approaches. This means that most efforts will have only a few tests of effects. Thus, at this point and for the foreseeable future, identification of developed and well-specified programs that have adequate empirical evidence is the preferred method for identifying a standard for practice.

What Standards to Use

Several lists of programs meant to reduce violence or related problems have been compiled, using varying methods and standards. Most are organized by benchmarks or criteria for inclusion (how programs are selected for review) and designation of level of confidence in the effectiveness of the listed programs (sorting into one or more levels of confidence that the program is effective). A few lists rate evaluations using multiple criteria to make an overall judgment, so that the rationale for a given program to be included or not is hard to discern. In some cases, ratings are provided, but the user is left to determine how these criteria might affect program value.[10] Noteworthy programs that have the soundest effects and are most preferred (usually highest designation) are those that used

---

[10] The correspondence of many programs' status across several listings can be found at http://www.blueprintsprograms.com/resources/Matrix_Criteria.pdf (accessed October 10, 2013).

the strongest evaluation methodology design (e.g., RCT or very carefully matched quasi-experimental).

Forming lists requires three major considerations. The first is how programs are identified for review. The second is determining the methodological standards for potential inclusion. The third issue is what is required to be listed as having beneficial effects and how differences in confidence about those effects are denoted. Systematic and assertive searching of the identifiable literature is needed to minimize bias and/or inconsistency in how and what programs are considered for evaluation and potential conclusion. For example, Blueprints has a policy of regularly surveying publication and online sources to identify reports as the source for evaluation of programs. Also, when program information is sent in, an additional search is done to identify all pertinent information. If potential evaluation literature is not systematically scanned, programs that are effective might be overlooked, and there is a bias for which programs are listed. This also creates confusion about what not being listed means. Similarly, there are problems when lists are developed from programs identified through organization-funded programs. These situations can create pressure to modify standards so programs that were otherwise not reviewable are included.

The Great Advantage of Random Assignment Trials for Determining Evaluation Effects

When program evaluations are identified, there is a second immediate consideration: whether the evaluation material available is of sufficient methodological quality to permit appropriate inference of effects (or lack thereof). Can determinations of effects be attributed to the intervention program and only to the program (Shadish et al., 2002)? This is the basis for an argument for random assignment as the standard or at least the much preferred design. Because the intervention and control conditions only differ by random determination of group assignment of a given person or unit of intervention, it greatly simplifies the ability to have confidence that any differences are due to intervention condition. All other methods have greater susceptibility to confound and biases that affect group assignment and therefore require more extensive and elaborate assurances that such biases did not occur. The advantages of random assignment have made it the standard for identification of effective approaches in many areas of public health and welfare.[11]

While random assignment has many advantages, there are numerous practical considerations that can constrain evaluations regarding reliance on it (Harris et al., 2006). Perhaps the most common is that participants

---

[11] For example, see Lachin et al., 1988, and Hedden et al., 2006.

or collaborating agencies will not agree to random assignment. If the fact that the intervention is not proven, which is the reason for the evaluation, is emphasized in presenting the rationale for random assignment, it can diminish enthusiasm for cooperation, or at least the fear of this can undercut preference for random assignment. In many instances interventions are developed out of interest of a group, setting, or agency (e.g., a school that sees the value in violence prevention or an agency that sponsors domestic violence advocacy services). In some cases, matching of a control to the voluntary intervention site is the most plausible design possible. However, such a compromise brings considerable decrement in confidence about the results obtained (Shadish et al., 2002). Concordantly, determining that results are not influenced by design bias (e.g., enthusiasm of comparison is lower than intervention) requires considerably more evidence. Thus, in many areas of health care the standard has become requiring random assignments for consideration for causal inference, showing that an intervention has or does not have direct and clear beneficial effects (Shadish et al., 2002; Hedden et al., 2006). Therefore, random assignment or very strong quasi-experimental design with accompanying statistical tests to ensure lack of bias in results are considered the minimal methodological requirements for an evaluation to provide evidence to determine intervention effects. This is the standard used in the Blueprints initiative.

Random Assignment Design Does Not Ensure
a Random Assignment Evaluation

While random assignment provides many strengths, including ease of interpretation, the conduct of a randomized assignment trial is vulnerable to many threats to maintaining the original characteristic of condition assignment only being due to random choice. For example, there can be differential attrition by condition. This means that those not participating, among those assigned and those leaving before the study is complete, may differ on important demographic, risk, or other characteristics that render the once-equitable groups not so for outcome comparison. If, for example, a violence prevention effort requires direct and open discussion of partner violence, it could result in those in the intervention condition leaving if they are engaged in more serious violence because of threat of arrest. This would render a difference between intervention and comparison condition that could explain any effects found, despite initial random assignment. Similarly, there could be loss—even if not different by condition—that could lessen confidence in results because the loss is related to how the program is expected to have effects or how these might be rendered. For example, it could be that all of the most at-risk families tend to move out of a parenting program aiming to reduce child abuse. While the average

scores may decrease in both groups over time, it would be difficult to rule out differences being due to this systematic change in the populations, even if they are similar across conditions. Other challenges of implementation, such as relatively low reach (low participation), uneven dosage (considerable variation in extent of exposure of those engaged in the intervention), or low fidelity of implementation (uncertainty of what was actually provided/experienced) can also render an evaluation with randomized design to no longer permit a sound interpretation of an intervention result as causal.

Features Other Than Collapse of Random Assignment Can Render an Evaluation Unusable

Other features of an evaluation, while not directly a failure of randomization or of adequate control for comparability for effects, can render a given evaluation not usable for determining a program's effects or lack of effects (Maxwell, 2004). Quite commonly, studies are not conducted at a large enough scale to be able to detect expected differences statistically (low power). This is particularly likely when the unit of assignment is not people, but groups of people (families within a neighborhood targeted for intervention), or organizations or social units in which people might be grouped (e.g., schools). In addition, the statistical analyses applied can be incorrect for the measures used, the units of assignment, or the expected effects. Most commonly, there are analyses of individuals even though random assignment was at the group level (e.g., shelter, marital couple, classroom, school).

While not affecting randomization or internal validity, additional criteria have been suggested to identify usable program progress. One such effort is that the intervention is tested with a sample representing the population of interest. Studies of convenience or opportunity samples may raise questions of generalizability—whether effects are meaningful for the populations affected by these violence issues. Another important consideration is that effects are on outcomes that are meaningful for the problem (e.g., if the goal is to reduce violence in marital conflict, the effects are on violence, not just stated attitudes about violence). A third consideration is that effects need to be accounted for by the processes or practices in the program thought to affect violence (Kazdin, 2007; MacKinnon et al., 2007). Thus, there is increasing interest in mediation analyses that demonstrate the program effects can be explained by theorized processes of effects, with some calling for discriminatory mediation analyses (Kazdin, 2007; MacKinnon et al., 2007).

Tough Standards?

The process of developing a program that can be tested is daunting. It involves undertaking evaluation with a design and scope that could permit determination of efficacy, implementation with fidelity and consistency, measurement of the key processes and outcomes with sensitivity and reliability, and maintaining randomization or very strong quasi-experimental comparability. These actions can require considerable expertise, resources, dedication, and artfulness. This has led some to suggest that softer standards or less insistence on set standards is in order. The intent is not to make it prohibitive to be recognized as a valid program and to ensure that programs developed without these capabilities are accessible by list users. Counterpositioning design adequacy for scientific judgment and the concomitant challenges seems to conflate two important but distinct considerations. While recognizing the challenges and constraints that affect ability to conduct such evaluations, it is also worth noting that these design and evaluation completion requirements are for basic evidence of effects—to provide the evaluations that could be used to make statements about programs that have promise or can be models for use. Eschewing standards will not lead to sound program choice guidance. If viewed as basic requisites for judging the use of scientific standards, then these challenging requirements may be seen, nonetheless, as necessary for developing a reliable, valid, and transparently rendered roster of programs that can be used for violence intervention. Another helpful step may be to identify areas of need for substantial efforts to conduct trials that can fill in gaps in knowledge in areas such as domestic violence, child abuse, and elder abuse (Tolan et al., 2006).

Scientifically Determined Standards for Determination of Grouping into Levels of Empirical Basis

Because the primary purpose of setting standards for inclusion on lists and compiling lists is to provide efficient guidance to those engaged in funding, practice, policy making, and administration (in addition to adequate methodological design in evaluation and maintenance of that quality throughout the evaluation), the benchmarking used to identify programs included on lists is another important consideration. The basis of different designations (e.g., promising, model, ready to go to scale, ineffective) need to readily understandable, reliably determined, and transparently applied. For example, one Blueprints designation is promising (requires at least one RCT or two quasi-experimental design studies meeting design quality requirements as summarized above), with significant immediate or longer term effects and no health-compromising effects. This designation means exactly what the category title says. These are programs that have been

reviewed and show promise as violence interventions, based on all relevant data. The second level is labeled model, to indicate these are programs that can be relied on for use. The requirements to be considered a model are (design quality requirements as summarized above) two RCTs or one RCT and one very strong quasi-experimental design evaluation, each with promising positive effects and no health-compromising outcomes. Also, effects must be sustained for at least 12 months after intervention on at least one outcome.

During the past 15 years, the Blueprints staff and advisory board have surveyed and evaluated approximately 1,000 programs (the procedures of the review and data collected and made available from the reviews is in Box II-1). Of those initially considered, about 150 have design standards that appeared adequate for full review. Approximately 39 of these have been designated promising and approximately 9 have been determined to be model programs. (These numbers are increasing, but in addition, as all programs are reviewed again periodically and as new pertinent evaluations are found, programs can change designation and/or be removed from the list.)

Noting that the 1,000 programs surveyed represent a small portion of the variety of efforts being used (and funded) for youth violence—and that most of those do not have evaluation quality to permit determination of effects using basic scientific standards of adequate design—it is still concerning that about 30 have met criteria for promising and only a dozen more have met the model criteria, the standard meant to convey readiness for use. This pattern highlights the extent to which inadequate attention to evaluation strengths and to replication of promising programs is constraining the ability to know whether most violence programs are having any effect.

Limits of List Making with Scientific Standards

The vast majority of violence programs in operation do not have evaluation information that could indicate effects. Some have suggested that lists are too constraining. Some have argued, and cannot be refuted, that there could be many programs that are beneficial, but have not been evaluated in a manner that makes them eligible for listing. However, this seems to argue for more careful and sound evaluation, not forgoing standards or obscuring what list inclusion means. The vast sums of money put toward violence prevention and its concrete importance are both powerful arguments for increasing attention to, funding for, and expectations of stronger evaluations and greater reliance on programs with evidence of effects.

A second limitation is that, to date, these review and listing efforts have been concentrated on youth violence perpetration, not on youth violence victimization or intimate partner violence, child abuse, or elder abuse

---

BOX II-1
Assertive Search Procedures and Program Information
Used by Blueprints for Healthy Development

Search Procedures

1. Systematically search for program evaluations, published and unpublished
2. Systematically review reports for evaluation methodology quality to be included for consideration
3. Those meeting study design quality standards to validly evaluate effects reviewed by independent advisory board
4. Individual programs with positive effects on meaningful outcome are certified as promising or model programs, depending on strength of evidence
5. Only model programs are considered eligible for widespread dissemination
6. Organize Program Summary and, for model programs, Program Implementation Guidelines Summary

Program Information Recorded for All Reviewed Programs

- Program name and description
- Developmental/behavioral outcomes
- Risk/protective factors targeted
- Contact information/program support
- Target population characteristics
- Program effectiveness (effect size)
- Target domain: Individual, family, school, community

(Tolan et al., 2006). Also, for the most part the reviewed programs are focused on the United States and Western Europe. Thus, perhaps the most critical concern about list making is the lack of such efforts for these other forms of violence and for a broader set of populations. For example, there are few quasi-experimental and even fewer randomized controlled trials on IPV, child abuse, and elder abuse. Those with the strongest methods point to relationship-based approaches, particularly for situations with less than the most extreme threat of harm (Tolan et al., 2006). Yet, these are preliminary and suggestive results at best. Moreover, the viability of such approaches is not a simple issue. This approach is not the focus of the majority of funding for such interventions. More sound evaluations are critical for improving the ability to affect partner, child, and elder abuse. This understanding of "what works" may provide much-needed basic direction, like it has in understanding youth violence.

Another major limitation of list making is programs that are methodologically strong, but may not have good evaluations of fit to particular

populations or communities, or problems may appear on the list. They may, by virtue of inclusion, take on the primatur of benefits for groups about which they do not have evidence of effects. Although more and more programs are considering these issues in design and in evaluation, most have been tested with specific (actually often unspecified) populations and with relatively weak tests of impact by gender, age group, economic levels, ethnic groups, or community type. Few culture-based or culture-specific programs have had the quality of evaluation to permit inference about effects. This limit is also applicable in regard to international differences in needs and resources. In addition to concentrated efforts to improve evaluation confidence, there does seem to be value in preferring programs with adequate evaluation and evidence of positive effects over those without such evidence.

A third criticism raised about the development of lists with the scientific standards used in other areas of health care is that many established efforts would need to be dropped in favor of efforts that may not have community support. They also may require reorganizing violence prevention efforts. Multiple practical and financial considerations would prompt this criticism. However, it is hard to see this argument prevailing if it is recognized that the current accepted efforts have no sound evidence of making a difference. Programs are being supported for reasons other than the effects they produce. There are multiple examples of efforts that while thought to be valuable, even by the affected communities, were in fact ineffective (no positive effects) and may even increase risk (Elliott and Tolan, 1999).

Ongoing Issues for Blueprints and Other List Approaches

Although Blueprints' list formation efforts and others like it can provide important direction and information toward the goal of effective violence reduction, there are emerging and ongoing issues related to inclusion criteria, review criteria, and determination of preferability. For example, one issue is how replication is determined (Valentine et al., 2011). To replicate a program, how much can content or implementation vary for the results to be considered rendered from the same program? If a program focused on parent training adds a few social-cognitive sessions to promote youth self-control, is this considered a variation or a different program? How important is variation in mode of exposure? If offered in person, is a program equitable to the same approach and activities offered through the Internet?

Another set of issues relates to how effects are judged (Aos et al., 2004). A key criticism of the replication approach used in Blueprints is how statistical significance should be considered (Valentine et al., 2011). A related issue is how the size of effect should be considered, meaning how large of an effect is valued whether meeting standards of statistical significance or

not. Some have argued that effects need to be of a certain size to be meaningful. Even if statistical significance is found, it should not be enough to judge a program as being effective. One step in that direction has been to calculate the benefits of programs in reduced costs (e.g., criminal justice, education, employment) that can be attributed to reducing violence (Aos et al., 2004). This method compares the cost of the program to the benefits based on the effects over time. The approach has particular appeal because it translates effect sizes into economic calculations that show return on investment for different programs. For example, funding that was dedicated to building more youth correctional facilities was redirected by the state of Washington, based on the cost-effectiveness estimates of increased reliance on empirically tested early intervention and youth violence prevention in that state (Drake et al., 2009).

As there is a growing body of adequate evaluations of some programs, the question of how meta-analytic methods, which measure effects as the average across the pertinent evaluations, should be considered (Valentine et al., 2011). At present, even the most evaluated programs have a relatively small number of evaluations. As noted by Valentine et al. (2011), meta-analysis has many strengths for testing for replication or consistency of findings across studies and for identifying robust estimates of effect size of a program or approach. This is a different set of criteria than the current Blueprints approach, which focuses on replication through independent studies, each with statistical significance. The relative advantages and limitations of the meta-analytic approach versus the independent replication approach are discussed extensively in Valentine et al. (2011) and a related commentary by the Advisory Board of Blueprints for Violence Prevention (2011). What seems to be agreed upon is the value of supporting evaluation of programs and multiple evaluations so that meta-analytic approaches can be applied to provide robust estimates of effects. Until that situation is reached, however, interim standards may be needed when there is a small number of sound evaluations.

Another aspect of violence intervention evaluation that can be vexing is how group-based interventions should be considered. Because such interventions involve randomizing large units and therefore often larger costs and administration requirements, the scale needed for valid randomized controlled trials can be daunting. To ensure inclusion of such interventions, some consideration of these factors seems warranted. How to do so without compromising the scientific standards—which do not vary based on unit of evaluation/assignment—is an ongoing concern with likely evolving standards. A related concern is how those programs meant to change the ongoing developmental environment should be considered (e.g., change procedures used in school for how teachers manage students' misbehavior). These programs are not simply applied to a group of youth and then

followed for a long-term effect on that cohort, but are meant to change how ensuing cohorts are affected. The question of how to evaluate effects on subsequent cohorts (e.g., do practices continue?) and the question of how to measure the "end" of the program are both important considerations, particularly as such efforts become more common (e.g., legal changes in how domestic violence is to be prosecuted).

Similar issues arise in considering program delivery efforts such as Communities That Care (Hawkins et al., 2002), which are not specific prevention programs, but instead are focused on how communities organize to implement prevention. Thus, the effort is not a specific program with a particular group, but a method of engaging community leaders in use of evidence-based programs that fit the identified risk and protective factors of that community. As with efforts to change organization in schools, these approaches are indirect in the sense of changing the operational setting, which are then thought to change the conditions for youth development. They are also similar in having less clarity about when the intervention ends. These advances in sophistication and breadth of approaches to prevention raise new challenges for any evaluation of approaches that is meant to differentiate "what works" from what does not and what is not properly evaluated. Thus, these are important and welcome challenges for the Blueprints and similar approaches to development of lists.

Program Lists and Moving Forward in Violence Intervention

This report has focused on Blueprints as an exemplary approach to violence prevention because it is a transparent, sound, and reliable standard that has many advantages over other approaches to list development and other efficacious approaches to identifying preferable programs and practices. However, as noted, there is a need for much more evaluation, including multiple evaluations of most promising approaches and model programs. In addition, a key need is to align listing efforts so that consumers—whether funders, administrators, policy formulators, or state and local agencies and groups implementing violence prevention—can readily understand the basis for listing a given program in a given category (e.g., promising, model, unproven, negative effects). There are currently considerable impediments in using the sounder programs because lists have varying quality of standards. Another factor is the varying extent to which listing occurs only if the programs have sound evidence of positive impact. Box II-2 lists a categorization schema developed among several agencies and groups. This effort to standardize list criteria and the terms for different levels of strength of evidence and reliability of use, if adopted in violence prevention across agencies and countries, would improve reliance

---

BOX II-2
Suggested Schema for Hierarchical Program
Classification Across Lists

1. Model: Meets all standards
2. Effective: RCT replication not independent.
3. Promising: Q-E or RCT, no replication
4. Inconclusive: Contradictory findings or non-sustainable effects
5. Ineffective: Meets all standards, but with no statistically significant effects
6. Harmful: Meets all standards, but with negative main effects or serious side effects
7. Insufficient Evidence: All others

NOTE: Q-E = quasi-experimental; RCT = randomized controlled trial.
SOURCE: Adapted from review of classification systems for program effectiveness ratings (see http://www.colorado.edu/cspv/blueprints/ratings.html).

---

on empirically tested programs (perhaps including the support for better evaluations).

Similarly, greater integration of usability concerns into program development and evaluation designs is an important step in closing the gap between what has been evaluated and what is readily useful for implementation at the community level. Box II-3 provides a summary of the characteristics for an "ideal" evidence-based program. Continuing to pursue this ideal and to promulgate sound lists can be an important contributor to effective violence intervention.

BOX II-3
The "Ideal" Evidence-Based Program

1. Addresses major risk/protection factors that can be changed and substantially affect problem
2. Easy to implement with fidelity
3. Rationale for and methods of services/treatments are consistent with the values of those who will implement
4. Keyed to easily identified problems
5. Inexpensive or positive cost/benefit ratios
6. Can influence many lives or have life-saving types of effects on some lives

SOURCE: Adapted from Shadish et al., 1991.