

The Robert Wood Johnson Foundation sponsors the Fighting Back initiative, in which communities across the country receive funding to combat alcohol and drug use. The evaluators of this initiative have tried to present an evaluation that is responsive to both national policy agendas and local community contextual differences, and thus offer important insight for both levels.

The View from Main Street and the View from 40,000 Feet: Can a National Evaluation Understand Local Communities?

Leonard Saxe, Elizabeth Tighe

As evaluators, we take it as a given that programs need data to determine whether their efforts have been effective. Local communities are taking greater control and responsibility for program design and implementation. Therefore, evaluators need to achieve a better balance between efforts that develop valid conclusions that can be replicated across situations and time and conclusions that are of immediate import to sustain or maintain the program. For policy-making, our goal in evaluating local programs remains a synthesis of results to distill underlying principles about behavior and community change. These can then be applied to other communities. At the same time, involving community leaders can increase the sensitivity of the evaluation to local issues, and thus make the evaluation more useful. Our evaluation of Fighting Back exemplifies such an effort.

This chapter describes an approach to incorporating two models of evaluation: an accountability model and an empowerment model. Our focus is how both to stand back from a program and view it objectively, while also being on the ground and helping to collect data that can be used to develop and support the program implementation. The scientific accountability model of evaluation has its roots in efforts from the Great Society era to make social programs accountable. A model of evaluation based on scientific methods evolved whereby social programs were viewed as "treatments" that were compared over time or to comparison groups (Campbell, 1969; Saxe and Fine, 1981). Although there has been criticism of the model, it does address the questions of external audiences

concerning the outcome of policies. This model has always existed in tension with efforts to empower those with responsibility for program development and implementation (Fetterman, 1996; Patton, 1978; Stake, 1975). The tension is increasingly evident as accountability is enforced and, at the same time, decision-making reverts to those with direct program responsibility.

Devolution of responsibility to the local level has been key to social innovation in government and social services since the 1970s. Programs that were formerly designed and managed by the federal government are now bloc-granted to states and local agencies. Local communities have a new set of responsibilities; no longer merely held accountable for managing programs, they are now responsible for how the program is designed and implemented. Thus, for example, substance abuse policy has increasingly focused on the development of community-based programs to reduce demand for drugs and to implement comprehensive identification and treatment services (Office of National Drug Control Policy, 1998; Winick and Larson, 1997). Community representatives involved in such programs need evaluative feedback to manage their initiatives. Yet community representatives are not the only stakeholders; in particular, policymakers need to know if these types of programs are effective. Serving both sets of stakeholders is a significant challenge; however, we have discovered that their needs can be compatible. For evaluation to be effective, it needs to satisfy the mutual interests of communities and national sponsors.

To illustrate the synergy between local and national perspectives, we focus here on the Fighting Back program of the Robert Wood Johnson Foundation (RWJF). Fighting Back is the first national demonstration of communitywide substance abuse prevention (Jellinek and Hearn, 1991; Spickard, Dixon, and Sarver, 1994). Communities designed site-specific approaches to reduce the demand for alcohol and other drugs (AOD) and related harms. This was accomplished by developing coalitions among public agencies, private organizations, and citizens' groups. These groups were charged with developing and implementing a comprehensive plan to deploy AOD programs across the continuum of care in their community, including public awareness, prevention, early identification, treatment, and aftercare.

Local control over program design and implementation was central to the Fighting Back initiative. The theory behind Fighting Back employed several assumptions about communities. First, each community was assumed to have unique characteristics that affected the nature of its substance abuse problem. Second, effective programming was believed to be more likely when grassroots leaders increased communication with managers of programs (including health care, social services, and police) than when external experts, unfamiliar with the local context, imposed their solutions. Citizens needed to be involved in the development of solutions in order to change communitywide attitudes and behavior about substance use and misuse. In addition, emphasis was placed on local control because external solutions had focused almost exclusively on supply reduction (reducing the availability of drugs) rather than on demand

reduction (reducing the desire for drugs). At the time Fighting Back was conceived, the nation was witnessing high levels of inner-city crime associated with substance abuse—in particular, violence associated with the use and sale of crack cocaine (Jellinek and Hearn, 1991; Musto, 1992). Reducing demand was perceived as key, but no models existed to reduce both supply and demand at the community level.

Although the demonstration was designed to test a new model of prevention, the program was developed in communities that were already experimenting with ways to deal with substance abuse. Not uncommon for evaluation studies, the difference in focus between the policymakers and the community leaders created tension that evaluators often felt. Tension between the policymakers' need for causal information and local stakeholders' concern for supportive feedback underlay the evaluation of Fighting Back. These two concerns had to be balanced by the evaluation design.

Although Fighting Back is a national initiative funded by a private philanthropy, it has many government parallels. Among the substance abuse prevention programs, the Community Partnership Programs of the Center for Substance Abuse Prevention (1996) are the most prominent and probably most similar to Fighting Back. Through this initiative, over 250 partnerships were awarded up to five years of funding. Typically partnerships served entire counties, although some served smaller areas. The ultimate goal of these programs was to change the community in such a way that everyone would feel responsibility toward, and be involved in, preventing substance abuse. Other drug prevention programs supported by the national government include D.A.R.E., Drug Free Communities, and various efforts to reduce crime and other harms associated with drug use (Winick and Larson, 1997).

Validity Versus Utility

The evaluation of locally controlled programs, such as the Fighting Back demonstration, requires integrating models focused on validity with those focused on utility. Shadish, Cook, and Leviton (1991) characterize the differences between these models in terms of the experimenting society (Campbell, 1969, 1988), explanatory theorists (Cronbach, 1982; Weiss, 1977), and stakeholder models (Patton, 1978, 1997; Stake, 1975, 1980; Wholey, 1983). Each of these perspectives has strengths and weaknesses. Although evaluators recognize that greater consideration should be given to a model of evaluation that combines the three perspectives, they rarely do so. Typically evaluators adopt a single perspective, and they emphasize either validity or utility. The evaluation of Fighting Back tried to maximize both validity and utility and to play to the strengths across models rather than the differences between them. The emphasis is on truth and objectivity in evaluation rather than advocacy (Scriven, 1997).

Well-designed experiments maximizing internal validity have been important to the development of alcohol, tobacco, and other drugs (ATOD) programs. For example, in the mid-1980s, amid growing concern about the use

of alcohol, drugs, and tobacco by youth, Botvin and colleagues developed the Life Skills Program designed to strengthen students' abilities to resist peer and social influences to use ATOD (Botvin, Baker, Filazzola and Botvin, 1990). In 1985 dozens of schools in New York State were randomly assigned either to receive or not receive this program. Six years later, students who had participated in the Life Skills Program showed significantly lower rates of substance use than comparison group counterparts (Botvin and others, 1995). These studies appear to have good internal validity, but some threats still exist because of the context.

The Life Skills Program cannot occur without all of those involved (students, teachers, school administrators) knowing about its desired consequences. Students reported lower rates of substance use on standard questionnaires, but the specific effects of the Life Skills Program cannot be separated from the additional effects of fifteen hour-long education sessions for teachers and other adults. As Campbell (1971) noted, the added variation introduced by the social context will inevitably be confounded with the intervention, making it difficult to interpret observed outcomes, regardless of randomization.

Given that a completely unbiased social experiment appears infeasible, alternative models, such as the explanatory and stakeholder models, seek to emphasize utility in evaluative research. These models emphasize the complexity of social problems in their context. Rather than studying simple causal relationships between a limited number of controllable variables, evaluation should model the complexity of social problems as they exist in the social world. This includes gaining better understanding of why a program might yield positive results in some contexts and no results or reverse effects in others by explaining higher-order interactions, emphasizing external validity over internal validity (Cronbach, 1982; Chen and Rossi, 1983; Weiss, 1978). It also includes evaluations that are of immediate and direct use to those involved in implementing the programs (Patton, 1978; Stake, 1975, 1980; Wholey, 1983). For example, despite possible problems with the interpretation of causality associated with the Life Skills Program, the intervention appears to provide many benefits to the students who participated (Botvin and others, 1995; Hawkins, Catalano, and Miller, 1992; Tobler, 1986; Tobler and Stratton, 1997). We may not understand why the program is effective, but the data make clear that whatever is done is useful. That is, under certain conditions, positive effects are obtained, regardless of what underlying construct might have been manipulated.

This emphasis on the utility of evaluative research regardless of the level of experimental rigor is best represented in the stakeholder models of evaluation. The pinnacle of this evaluation perspective is perhaps represented in current discussions of empowerment evaluation (Fetterman, Kaftarian, and Wandersman, 1996). Fetterman (1996) describes empowerment evaluation as having its roots in community psychology. Thus, it is understandable that this perspective has gained prominence as social programs become more locally or community based. Empowerment, however, described as facilitating self-

determination through evaluation, is similar to the more utilitarian models of the 1970s (Patton, 1978; Stake, 1975, 1980; Wholey, 1983). Called responsiveness models or management-centered models, their focus has been on making evaluation useful.

Although evaluation research needs to be useful, to the extent that evaluation becomes confounded with program implementation, the greater are the threats to the validity of the evaluation process (Scriven, 1997). Evaluation does not have to be valid in order for it to be useful, or at least perceived to be useful. Thus, for example, Beveridge and others (forthcoming) have described how data from the National Household Survey on Drug Abuse were used to show that the Miami Coalition Against Drug Abuse had reduced substance use by 50 percent. The finding was touted in advertisements and prompted a visit by President Clinton to Miami to declare victory in the war against drugs. Unfortunately, the finding was artifactual. It was most likely the result of Hurricane Andrew. The hurricane devastated parts of the target area and made it impossible to sample similar households after the storm.

Fighting Back

Fighting Back has contended with the balance issue throughout its history: how to hold sites accountable to objective standards of program effectiveness, while being supportive of and sensitive to their ongoing program development. The national evaluation of this program was implemented primarily as an outcome evaluation: the ultimate question to be answered was whether rates of AOD use, and harms from use, declined significantly in the communities in which the program was implemented (Rindskopf and Saxe, 1998; Saxe and others, 1997). We employed a quasi-experimental design to examine twelve of the Fighting Back communities, along with thirty matched comparison communities. We developed standardized measures of use and harm that could be used and analyzed across all communities. These included both a random-digit-dial survey and archival indicators, such as crime data and mortality. In addition to these standardized outcome measures, the evaluation included in-depth ethnographic studies in several of the communities and a management information system implemented in all twelve Fighting Back communities to track program implementation.

Fighting Back then moved into a new phase with additional funding for eight of the communities participating in the project. At this point, assessment specific to localities and deemed useful became increasingly important in the evaluation. Communities needed local data for their own planning. Future funding became dependent on a community's ability to assess its progress in achieving reductions in AOD use and harm. Toward this end, the funder required sites to develop community reporting systems, modeled after community epidemiological working groups (CEWG). The data were intended to assess community progress in achieving AOD and harm reductions, but also for local planning.

The new phase of Fighting Back has altered the role and use of data. Not only do the sites and evaluators now have different interests, but also the evaluators and program managers have increasingly different concerns. The evaluators are concerned with drawing valid inferences about the program, while the program managers must ensure that the sites do well. Sites now seek to report local outcome levels. To address this problem, both the evaluators and program managers must be attuned to motives that can influence data collection and possibly lead to confounders. The need for the national evaluation became more directly involved in the local evaluation process. Some of the local data are a subset of the national evaluation data. Other locally specific information is not collected by the national evaluation and must be acquired and analyzed locally. Portions of these data are necessary for the national evaluation but need to be collected by the local community in collaboration with national evaluators. Thus, local assessment from the national evaluation perspective takes several forms, and each form has different implications for the evaluation and for the participating communities.

National Evaluation Data. Although collected in a common form across sites, the national data disaggregated to individual communities have potentially important utility for planning. Communities need assistance in gaining access to these data, interpreting the results, and using them appropriately to identify and respond to their communities' substance abuse problems. The interpretation and appropriate-use issues are significant, especially in the context of national-level survey samples that require different weighting schemes when examined at local levels. Policymakers have perhaps overemphasized "scorecarding" substance abuse rates; often scorecards ignore the context of a community and the extraneous factors that affect AOD use and attitudes. Also, random changes in substance use rates make them complex to use. Survey and indicator findings become inappropriate to compare. Finally, most of the national measures became available only after a considerable lag. The national data can provide rapid feedback only under limited circumstances. Nevertheless, the evaluators were in a position to help communities use the available survey and indicator data in their planning.

Locally Collected Data. The national evaluation was designed to test the concept that communitywide strategies could address the prevention and treatment of substance abuse. For this reason, we sought measures of the use and harm caused by substance abuse that could be assessed across the communities. Each community may offer unique and important data—unique, at least, in terms of how they are collected. School surveys of drug and alcohol attitudes and behavior can illustrate the issues. For both practical and legal reasons, it is infeasible for the national evaluation to collect its own data through school surveys (Hunt, 1998). Most communities, however, conduct school surveys or would be interested in doing so. The data from these surveys can provide essential information about the impact of Fighting Back. The national evaluators therefore supported communities in conducting their own surveys through technical assistance and material resources, if needed. Because the surveys are locally controlled, it is likely that each will use slightly different instru-

ments and designs. The evaluators can use the data and deal with the differences in instruments by applying meta-analytic techniques. The school survey situation may be unique; in other cases—for example, data about school dropouts or criminal activity—locally collected data would not be used directly by the national evaluation. It would be important, though, for the national evaluation to be involved in the development of such studies to ensure quality and avoid conflicts with other data collection activities. If successful, these efforts could provide important and timely information about the impact of Fighting Back.

Evaluation of Specific Initiatives. The Fighting Back program concept does not specify which interventions particular communities will adopt. Interventions may change frequently, in both their specifics and the role of Fighting Back. It is important and useful for communities to evaluate these interventions. The national evaluation team can help to ensure that these studies are likely to yield valid and useful information.

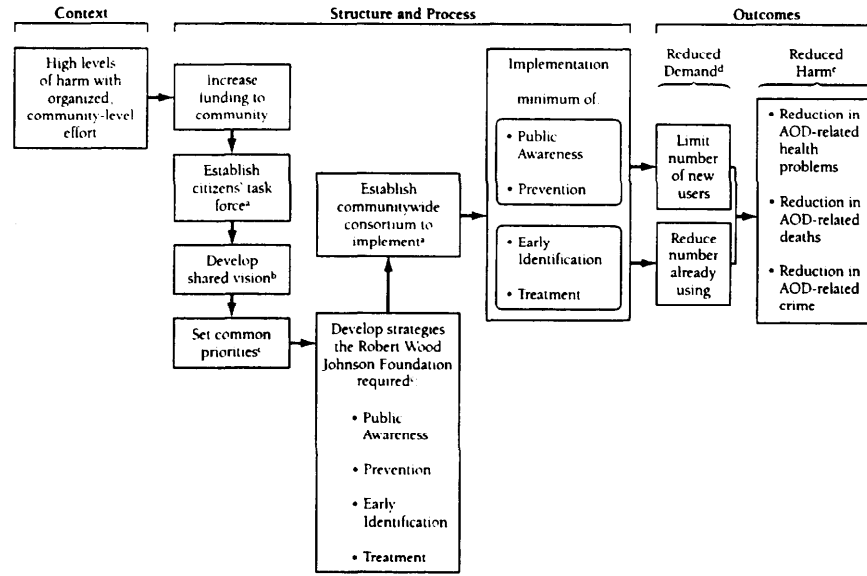
The new role for evaluation data for this phase of Fighting Back has several implications. Most important, the need has become acute to maintain independence between the evaluation and program functions. If they are constrained by a need to “make the program look good,” the evaluators cannot conduct an objective assessment and provide evidence that will convince skeptical policymakers of the value of Fighting Back. Our position therefore is that the evaluators need to be responsible for working with communities on local assessment issues. Where appropriate, the evaluators would provide data, analyses, and technical assistance.

We have also sought to infuse locally driven data collection efforts with the type of objective standards applied at the national level. We worked with communities' questions about local data to focus on the three key principles: validity, utility, and feasibility. *Validity* means these data should measure what the underlying constructs intend. *Utility* means the data must also be useful to one or more of the stakeholder groups. Although they cannot always determine this a priori, both evaluators and community members should think through how they might use the findings. Finally, the data must be feasible to collect. *Feasibility* encompasses two considerations: whether the measure is likely to be sensitive to detecting change reliably (that is, effect sizes are likely to show statistically significant change) and whether the data can actually be collected and analyzed in the time frames proposed. Focusing discussions about data at the local level on these three key features of the evaluation process has helped guide local data collection. It has also helped to inform the national evaluation team about issues that are most relevant at the local level.

Logic Models and Indicators at the National and Local Levels

The underlying model for the Fighting Back program guided data collection at the national level (see Figure 5. 1). A community context defined by high levels of AOD use and harm leads to the motivation to develop comprehensive,

Figure 5.1. Fighting Back Logic Model



^aFrom *Call for Proposals* (p. 6) (parents, clergy, tenant groups, business and commercial leaders, health professionals, school officials, judges, police, and others).

^bFrom Jellinek and Hearn (1991, p. 81). Examples of shared vision might be whether to deal with substance abuse directly or with underlying causes.

^cFrom *Call for Proposals* (p. 5) and Jellinek and Hearn (1991, p. 80). Priorities might be to target children, young adults, parents, or inpatients or outpatients, for example. If shared vision is to focus on underlying causes, then the effort might target the environment and other community needs.

^dFrom *Call for Proposals* (pp. 5–6) and Jellinek and Hearn (1991, p. 79).

^eFrom *Call for Proposals* (p. 5).

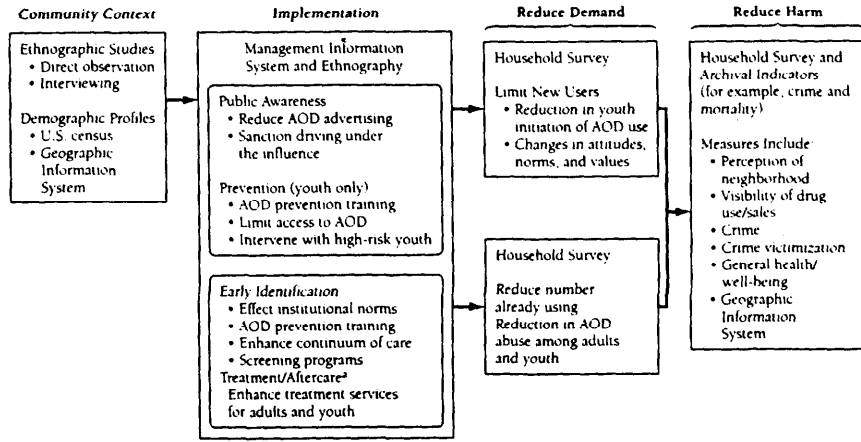
communitywide strategies to reduce demand for AOD, and thus reduce use and harm. Figure 5.2 focuses on the elements of this model that are central to the national data collection. Demographic data were used to develop a comprehensive picture of each community context. These were embellished with in-depth ethnographic studies in six of the Fighting Back communities. Further information about community context can be derived from the household survey, which includes perceptions of neighborhood, crime, attitudes and norms about drug use, and visibility of use and sales. These data, however, were not available prior to our initial involvement in 1995. Thus, they can be used only to assess community context after implementation. Data on reduced demand, in terms of the rates of new users and existing users, can also be inferred from the household survey, as can several measures of harm. The survey data should converge with other measures of harm indicated by standardized measures such as crime and mortality.

At the local level, data collection and analysis developed in a similar manner, but without the focus on standardized measures that could be uniformly collected across many sites. Sites focused on the program activities and goals

that were most likely to yield significant change by the next refunding period. Examples of the types of data that communities chose to examine and the relationships of these data to the underlying model of the program are depicted in Figures 5.3 and 5.4.

Figure 5.3 depicts a logic model and indicators for a community whose local data collection mirrors the national evaluation. It proposes to use data from the National Household Survey and from schools for indicators of reduced demand. For indicators of harm reduction, it proposes police arrest data for those arrests coded as AOD related as measures of harm. The community also

Figure 5.2. National Evaluation Measurement Model



*The household survey includes assessments on treatment and aftercare.

Figure 5.3. Measurement Model for a Community That Focuses Primarily on Ultimate Outcomes

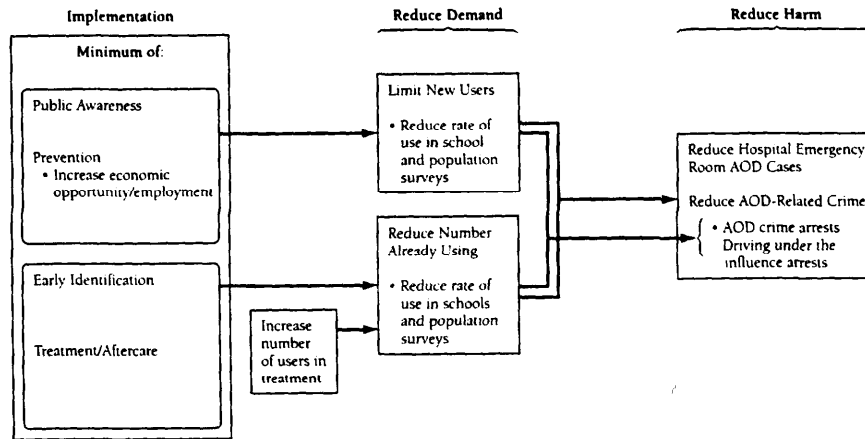
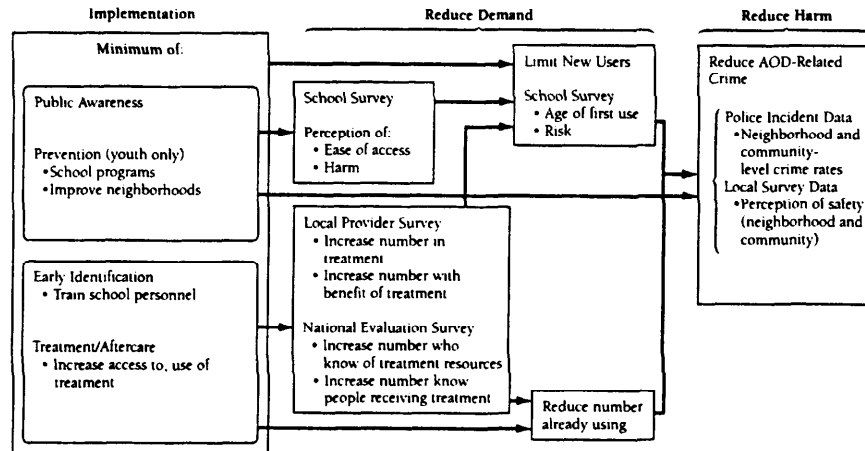


Figure 5.4. Measurement Model for a Community That Combines Ultimate and Intermediate Outcomes



proposes that these activities will result in a greater number of people receiving treatment. This is depicted as a mediating variable between program implementation and reduced rates of people using AOD. Several different sources are proposed to estimate the number of people in the community receiving treatment, including reports of state-funded treatment slots, raw data from drug court records, and possibly estimates derived from the National Household Survey. This community also proposes the unemployment rate as a marker of success, in that increasing economic opportunities for residents should prevent substance abuse. Although employment is viewed as an element of prevention implementation, the community may choose to narrow the measurement of employment rates to those in treatment. In that case, the indicator might be better characterized as a positive outcome of treatment rather than a measure of prevention.

In the other example of local data (see Figure 5.4), the community proposes a greater number of measures that represent the mediating variables between program implementation and the ultimate outcomes of reductions in demand and harm. There is also a greater mix of data collected both locally and by the national evaluation. Prevention activities occur for the most part in schools, and school surveys are proposed to obtain the primary effectiveness data. School programs are expected to change youth norms regarding use, in particular their attitudes about perceived harm and availability. Changes in attitudes should result in reduced rates of use, as well as reduced rates of youth at risk for problem AOD use. This community's overall strategy also includes an environmental strategy focused on improving deteriorating neighborhoods. The effectiveness of these efforts will be reflected in rates of crime and perceived safety. Crime rates will be based on police incident data, and perceived safety will be based on a local survey conducted by those involved in the

neighborhood intervention. These data should converge with household survey data on crime victimization and perceived safety in the neighborhood. The community also proposes to measure efforts at early identification and referral, as indicated by the number of school personnel trained. Increased access to and utilization of treatment services should be reflected in increased rates of people in treatment, which will be assessed based on a one-day sample of treatment providers' records. Although the site does not address the ultimate outcomes of communitywide decreases in AOD use and harm, the sources of data represented in this model should reflect steps toward these ultimate outcomes.

Analysis Across Local Assessment Models

These examples represent various compromises on validity to obtain maximum utility. Ideally, the national evaluation team will integrate local assessment models into the test of the underlying program construct, perhaps using meta-analytic techniques. Such a synthesis cannot be accomplished, however, without first taking into account limitations to the conclusions that can be drawn from locally collected data. Some elements of local evaluation designs may not be feasible. The strengths and weaknesses of the local data plans are best described in terms of the three key principles of validity, utility, and feasibility.

Validity. The most common problem across local data plans is a lack of fit between the program designs and the predicted outcomes. The lack of fit is an internal validity concern. Randomization was not possible at the local level, and in very few instances can comparison communities be studied. With the lack of randomization and control groups at the local level, issues of conclusion validity become a greater concern. And there is substantial variability in the degree to which the measures that have been selected reflect the program goals and activities.

Difficulties in operationalizing of constructs are shown in two areas: the ambiguous relationships among stated goals, activities, and predicted outcomes and the tendency to identify "measures" that refer to program activities rather than outcomes. For example, one community proposes to "expand treatment." Its measures of treatment expansion include rates of AOD-related hospital emergency cases, the number of health care workers trained in identification and referral, participation in aftercare services, and rates of AOD-exposed births. Each of these measures is unclear as a measure of treatment expansion. A more direct measure of whether the community has succeeded in expanding treatment would be the number of treatment slots available and whether these slots are filled by patients referred from the systems with which the community is working. Rather than focusing on hospital emergency cases, the goal of the hospital training is to get more people into treatment. Thus, a measure of the treatment services that are available, the number of people in treatment, and from where they have been referred might better reflect success in expanding treatment.

This is perhaps an extreme example of a situation in which there is a disjunction between construct and measurement. The measures in this case might still be useful, but for purposes other than the intent to measure treatment expansion. In many cases the measures reflect subgoals, or steps toward the ultimate goals. In many of the sites, however, there appears to be some confusion regarding how best to differentiate between subgoals and ultimate goals and how to incorporate the subgoals into the data plans. If the measures were operationalized in a manner that more directly corresponds to the goals or if the links to the goals were better explained, the utility of these data would be clearer.

Problems of valid measurement are also common. For example, another community proposed to examine the primary indicators of youth-related outcomes based on school survey data. For the past two years, the Fighting Back program has administered the school's substance use survey, with technical help from the national evaluators. For quite some time, the community has used the observed rates from these surveys to assess drug problems among youth. However, the survey design of these data has changed over time. Markedly different samples were drawn for each year, perhaps because of how informed consent was sought. Our involvement has led the community to modify both the way that the data are analyzed across time and the method for data collection. In this case, our focus on validity was useful to the local community and the school board.

Utility. As highlighted by the example, correspondence between measures and underlying constructs presents several problems in the interpretation and use of the data. One problem that affects the utility of the data is how proximal or distal the measures are to program activities. Another is the direction and specificity of expected effects.

Distal Versus Proximal Outcomes. The closer the unit of measure is to the actual activity, the more likely it is that the outcome will be sensitive to changes that occur as a function of that activity. Thus, communities that have more proximal outcomes are in a better position to demonstrate success than are communities that focus on distal outcomes. For example, one community proposed that its activities will result in significant changes in rates of youth and adult AOD use (as measured by surveys), rates of AOD-related crimes, and drunk driving. All of these reflect the ultimate outcomes originally expected of successful Fighting Back programs. However, they are not directly related to the specific activities occurring in the community. This community focuses on increasing AOD identification and referral to treatment at entry points into community systems, including schools, courts, hospitals, and police. Yet no measures are proposed at any of these entry points. Also, there is no measure of the community's primary goal: the identification and provision of services to substance abusers who come through the community systems.

In contrast, another community clearly distinguished between distal and proximal outcomes. Part of the plan for local data includes distal outcomes of program success that would be monitored but not used to evaluate progress toward these ultimate outcomes. Another part of the plan includes measures

that more directly reflect program activities in the areas of neighborhood revitalization, treatment utilization, and youth prevention. These outcomes indicate the steps toward achieving the ultimate outcomes. There are still some questions regarding the utility and validity of the measures, but in general the measures clearly are more closely linked to the program activities than they are in other sites.

Direction and Specificity of Predicted Effects. All communities propose either to increase or decrease outcome measures, but the predicted outcomes are often opposite to what would be expected as a result of their program activities. For example, several communities propose to decrease the rate of hospital emergency cases that are alcohol or drug related. Yet in many of these communities, significant resources are invested to train hospital staff to identify and refer cases. In the long term, if the program is successful (and this is what is needed to reduce harms from substance abuse), then indeed one might expect a decreasing rate of hospital cases that are substance abuse related. The assumption of current training programs, however, is that many cases are unidentified. Thus, in the short term (at least the next one to two years), the number of cases that are identified as alcohol or drug related should increase and be a marker of program success, not failure. Conversely, with this measure alone, an observed decrease in the rate of emergency cases that are alcohol or drug related is meaningless. This could occur because staff and others have not been effectively trained or because the problem has been reduced. Without measures of the problem (overall rates of substance abuse) as well as the effectiveness of the training (for example, evidence that staff members are not underreporting cases), it would be hard to discern what a decrease would mean.

Crime data represent another outcome measure that is predicted to be opposite in direction to what might be expected. Many communities propose to decrease the rate of crimes that are alcohol or drug related, yet work with area police units should result in an increased likelihood that cases are identified as AOD related. Conversely, a decreased rate of such crimes just as likely reflects police attitudes and policies instead of progress on the problem. It would be helpful if there were a way to document police enforcement and identification process or to converge reported crime data with crime victimization surveys and the visibility of drug sales in the neighborhood, obtainable from the National Evaluation Survey. One problem, however, is that the focus of crime problems has shifted from the general idea of crimes as community harms (reflected in homicide, assault, and burglary rates) to only AOD-related crimes. This shift narrows operationalization and causes inconsistent coding from site to site, making it difficult for communities to use the national survey for converging data. It is standard at the national level to examine Uniform Crime Reports (UCR) index crimes in combination with the Crime Victimization Survey, but if sites examine only AOD-coded crimes or incidents, convergent validity is nonexistent. Similar problems apply to other measures included in local data plans.

Feasibility. Feasibility reflects two issues: whether the data are possible to collect and interpret and whether the data are likely to yield significant change in the time period under investigation even if the site is completely successful at program implementation. Data the community would like to report cannot always be collected and analyzed, several communities propose to report data that do not exist, and other proposed data do not take a standard form that can be analyzed for change across time. Most communities have difficulty collecting treatment data, as well as rates of those in need who are receiving treatment. Some propose to estimate the rate based on the National Evaluation Household Survey data. Others prefer a census of all people in treatment and have rejected a survey estimate. Although some standard data exist about publicly funded treatment slots, none are available for private facilities. Rather than modifying their approaches, some communities have pursued primary data collection on their own—for example, by conducting one-day samples of those in treatment among all providers in their communities. Although these efforts will surely yield counts of people in treatment, the validity of the data and their reliability are yet to be determined. In addition, comparisons across time will be difficult since these sources of data are not available for periods before the intervention started.

The second problem is most obvious in instances in which the site proposes to report low-frequency outcomes, such as mortality rates or drug-exposed births. The national evaluation thoroughly investigated several sources of data, such as mortality and fatal vehicle crashes, and concluded that the rates of occurrence were too low for there to be sufficient power to detect significant changes across time (Beveridge and others, 1997). Thus, communities were dissuaded from including such measures, but this did not deter some who perceived that the data would be useful to examine even if they were not statistically reliable (such as rates of AOD-exposed births).

Strengths and Weaknesses of the Joint Evaluation Strategy

Conducting a community-level evaluation study is one of the most complex challenges that evaluators face (Connell, Kubisch, Schorr, and Weiss, 1995). The time frame for hypothesized change is long, the number of units that can be studied is small, the theoretical constructs are complex, and some of the most important effects may be obscured by other changes taking place. By developing a joint focus on outcomes common to many program communities and outcomes of interest to individual communities, we have tried to address many long-standing concerns about evaluation. The approach would seem to combine features of the experimental and stakeholder approaches. Nevertheless, it is difficult to determine whether the strategy is effective and whether it will improve the quality and usefulness of the evaluation.

The strengths of the joint strategy have been implicit in the discussion of how the national and local data plans were implemented. For national evalu-

ation purposes, the joint strategy permits the collection of cross-site data that can be easily compared to address the ultimate (distal) program goals. At the same time, the evaluators make their data available for local use, tailored to a particular site. They also work with local staff to develop measures sensitive to the unique situation of the community. To the extent that national and local data can meet the criteria of validity, utility, and feasibility, the findings can help to shape both national policy and local implementation.

Different Perspectives. In theory, the model incorporates the scientific requirements of evaluation, along with a utilization focus. In practice, however, the model may be difficult to implement because of the particular outcomes that are reported. A validity-driven study design applies stringent criteria to the measurement of program outcomes and is undoubtedly conservative in concluding that there are program effects (Rindskopf and Saxe, 1998). Although one can adjust the probability of a Type 1 versus Type 2 error, the scientific bias is toward accepting the null hypothesis. As evaluators have long noted, however, this creates a gulf between the perspectives of evaluators and program staff (Weiss, 1977). As an evaluator, one is inherently skeptical; as a program developer, one is enthusiastic about the intervention.

The national evaluation (mostly surveys of drug use and attitudes) has not yet demonstrated that Fighting Back communities achieved significant reductions in drug use (Beveridge and others, forthcoming; Rindskopf and Saxe, 1998; Saxe and others, 1997). This is particularly evident when sophisticated multilevel analytic models are used to contrast target communities with comparison communities. The findings contrast with the enthusiasm of local staff, who see daily evidence of their successful work. They see the individuals who are affected directly by their programs and are beneficiaries of changes in policy that Fighting Back brought about.

Some program staff have suggested that the national evaluation, because it views the communities from "40,000 feet," cannot see the positive impact of the program. Only when viewed from "Main Street" can one see how lives are transformed by the communitywide effort to reduce demand for illicit drugs. The problem is that having an impact on just those people one can see is not the goal of the program. In fact, evidence suggests that the visibility of drug sales and use is not an indicator of underlying patterns of use (Saxe and others, 1997). The program theory is that a communitywide effort is necessary to reduce communitywide levels of drug use and abuse. Thus, although the Main Street view is an important indicator of whether the program is making progress, the desired outcome goes beyond success with small groups of people.

The key question becomes how to scale up an intervention so that it can affect the overall rates of drug and alcohol problems. A delicate balance needs to be struck between different views of the achieved outcomes. The local view is necessary as a leading indicator, while the national data are the distal indicators sought by those who funded the program and those interested in the generalizability of the program strategy for other communities. A key issue is for both the national and local stakeholders to see the advantage of the others' perspective.

Between Ideal and Real. It is not sufficient to see the value of the other perspectives or to maximize the utility of data. Unfortunately, that which would be most useful is often invalid or infeasible to collect. All sources of data have limits to their validity. That is true for surveys of drug use (Beveridge and others, forthcoming) and local indicators of the harms associated with drug use. To be sure, in some cases it is possible to collect better data, but the cost may be prohibitive (as in the case of in-person household interviews). What seems important is to work toward balance among the three principles of validity, utility, and feasibility.

Balancing these principles involves strategic decisions about what data to collect and careful allocation of resources. With some demonstration programs, for example, it does not make sense to use an experimental design and extensive outcome-oriented measures. Particularly when the issue is whether the program idea is feasible, the research questions may not be suitable for this type of assessment. Such outcome data may not have utility in this context. Similarly, it may not be worthwhile for a local community to invest heavily in a data gathering system—even if the questions are outcome oriented—if the data cannot be gathered in a valid way. Thus, for example, it might be important to know if fewer crimes are committed as a result of new drug policies, but the collection of such data is dependent on the attitudes of both community residents (who report crimes) and police (who receive reports). Unless a valid system can be developed to collect such data, the investment is not worthwhile.

Few communities have the capacity or expertise required for primary data collection or analysis of secondary data sources. Although developing such capacity is probably useful, it can be done only over time and should probably be seen as part of a system of accountability and continuous quality improvement (Bickman and Noser, forthcoming; Wandersman and others, 1998). Where communities have collected primary data, the diversion of funds from program activities may not have been wise. Fighting Back communities have been encouraged to form local CEWGs so that the demands on local staff would be minimal, but this has yielded mixed levels of support. Each community was to have a full-time staff person who would coordinate the local data collection and analysis efforts. Although some Fighting Back communities have such a person, the task has proved difficult. The local data collectors have struggled to integrate their reporting on program success in a way that is acceptable to both their funders and the community group. Also, in sites without a full-time staff person that rely on local CEWGs, the community group acts more like an outside consultant with a narrow mandate. The group does not question the utility of the data or its relation to program activities; instead it tries to figure out how to collect data.

From the initial stages of our involvement with the evaluation of Fighting Back, we have been eager to provide sites with data. In part, we wanted to maximize the utility of the data; in addition, interaction with the sites is valu-

able to understand their perspectives on the problem. We conducted several sessions to feed back data and provided each site with a summary of data specific to their community. We were sensitive not to provide comparisons among sites and to provide contextual information that would aid understanding. These efforts were only partially successful. In retrospect, what was clear was that only a few sites had the capacity to use the data successfully and that substantial resources would be needed to assist the others to use data.

Sharing data with the sites has been useful but raises several concerns. The first has to do with a key change in *Fighting Back*: data may be used to make funding decisions. Providing data creates the potential for “dueling” evaluations whereby the national evaluators and local communities each have their own analyses—and reach their own conclusions. They may become motivated to look good so as to be refunded, and ours is “to find out the truth.” Moreover, a recompetition of funding based on results may be unworkable. The prevention and treatment of substance abuse is a multivariate problem. One community’s context may simply prevent changes of a size seen in other communities. The data elements cannot be treated in isolation from this context.

Although designed to empower communities to use evaluation, implementation of local data collection was not entirely up to communities’ free choice. Development of local data became a requirement of funding from the Robert Wood Johnson Foundation, and it is not yet clear whether all communities recognize the benefits of data about the nature of their problem and the outcomes. However, the exercise has other advantages. It may offer the same advantage as cost-benefit analysis (Office of Technology Assessment, 1980). The process of conceptualizing outcomes helps communities go beyond what they can see directly and enables them to think about planning in a different way. At the same time, it brings evaluators into close contact with the communities and requires that they think about the intervention in terms of very specific outcomes.

Healthy Tension. Kurt Lewin (1951), the progenitor of social psychology, posited that individuals hold multiple ideas in a system of tension. The relationship of national and local evaluation perspectives for *Fighting Back* is Lewinian, because we see the perspectives as being in productive tension. No single perspective provides the answers that will enable the development of more effective policy. Sometimes local and national perspectives will yield different conclusions. But these conclusions are not inherently in conflict; they may be the basis for improved thinking about the program.

To preserve a healthy tension, however, there must be fundamental agreement about the purpose of evaluation. For *Fighting Back*, these purposes are framed as principles for evaluation: validity, utility, and feasibility. Learning will be maximized if both community and national stakeholders are committed to finding feasible strategies to develop useful and valid information. The beneficiaries will be the citizens in each of the communities that have invested in the development of new programs and their evaluation.

References

- Beveridge, A., and others. *Monitoring Archival Indicators of Alcohol and Other Drug Harm*. Report prepared for the Robert Wood Johnson Foundation. Waltham, Mass.: Brandeis University, 1997.
- Beveridge, A., and others. "Survey Estimates of Drug Use Trends in Urban Communities: General Principles and Cautionary Examples." *Journal of Substance Use and Misuse*, Forthcoming.
- Bickman, L., and Noser, K. "Meeting the Challenges in the Delivery of Child and Adolescent Services in the Next Millennium: The Continuous Quality Improvement Approach." *Journal of Applied and Preventive Psychology*, forthcoming.
- Botvin, G. J. "Substance Abuse Prevention Research: Recent Developments and Future Directions." *Journal of School Health*, 1986, 56, 369-374.
- Botvin, G. J., and others. "A Psychological Approach to Smoking Prevention for Urban Black Youth." *Public Health Report*, 1989, 104 (6), 573-582.
- Botvin, G. J., Baker, E., Filazzola, A. D., and Botvin, E. M. "A Cognitive Behavioral Approach to Substance Abuse Prevention: One Year Follow-Up." *Addictive Behaviors*, 1990, 15, 47-63.
- Botvin, G. J., and others. "Smoking Prevention Among Urban Minority Youth: Assessing Effects on Outcome and Mediating Variables." *Health Psychology*, 1992, 11 (5), 290-299.
- Botvin, G. J., and others. "Long-Term Follow-up Results of a Randomized Drug Abuse Prevention Trial in a White, Middle-Class Population." *Journal of the American Medical Association*, 1995, 273 (14), 1106-1112.
- Campbell, D. T. "Reforms as Experiments." *American Psychologist*, 1969, 24, 409-429.
- Campbell, D. T. "Methods for the Experimenting Society." Paper presented at the meeting of the Eastern Psychological Association, New York, and at the meeting of the American Psychological Association, Washington, D.C., 1971.
- Campbell, D. T. *Methodology and Epistemology for Social Science: Selected Papers*. Ed. E. S. Overman. Chicago: University of Chicago Press, 1988.
- Center for Substance Abuse Prevention. *Fourth Annual Report of the National Evaluation of the Community Partnership Demonstration Grant Program*. Rockville, Md.: CSAP, U.S. Department of Health and Human Services, 1996.
- Chen, H., and Rossi, P. H. "Evaluating with Sense: The Theory-Driven Approach." *Evaluation Review*, 1983, 7, 283-302.
- Connell, J. P., Kubisch, A. C., Schorr, L. B., and Weiss, C. H. (eds.). *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Washington, D.C.: Aspen Institute, 1995.
- Cronbach, L. J. *Designing Evaluation of Educational and Social Programs*. San Francisco: Jossey-Bass, 1982.
- Fetterman, D. M. "Empowerment Evaluation: An Introduction to Theory and Practice." In D. M. Fetterman, S. J. Kaftarian, and A. Wandersman (eds.), *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. Thousand Oaks, Calif.: Sage, 1996.
- Fetterman, D. M., Kaftarian, S. J., and Wandersman, A. (eds.). *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. Thousand Oaks, Calif.: Sage, 1996.
- Hawkins, J. D., Catalano, R., and Miller, J. "Risk and Protective Factors for Alcohol and Other Drug Problems in Adolescence and Early Adulthood: Implications for Substance Abuse Prevention." *Psychological Bulletin*, 1992, 112 (1), 64-105.
- Hunt, M. *The New Know-Nothings: The Political Foes of the Scientific Study of Human Nature*. New Brunswick, N.J.: Transaction Publishers, 1998.
- Jellinek, P. S., and Hearn, R. P. "Fighting Drug Abuse at the Local Level: Can Communities Consolidate Their Resources into a Single System of Prevention, Treatment, and Aftercare?" *Issues in Science and Technology*, 1991, 7 (4), 78-84.

- Lewin, K. *Field Theory in Social Sciences: Selected Theoretical Papers*. Ed. D. Cartwright. New York: Harpers, 1951.
- Musto, D. F. "Historical Perspectives on Alcohol and Drug Abuse." In J. H. Lowinson, P. Ruiz, R. B. Millman, and J. G. Langrod (eds.), *Substance Abuse: A Comprehensive Textbook*. (2nd ed.) Baltimore, Md.: Williams and Wilkins, 1992.
- Office of National Drug Control Policy. *The National Drug Control Strategy, 1998: A Ten Year Plan*. Washington, D.C.: Executive Office of the President of the United States, 1998.
- Office of Technology Assessment. "The Implications of Cost-Effectiveness Analysis of Medical Technology," 1980.
- Patton, M. Q. *Utilization-Focused Evaluation*. Thousand Oaks, Calif.: Sage, 1978.
- Patton, M. Q. *Utilization-Focused Evaluation: The New Century Text*. (3rd ed.) Thousand Oaks, Calif.: Sage, 1997.
- Rindskopf, D., and Saxe, L. "Zero Effects in Substance Abuse Programs: Avoiding False Positives and False Negatives in the Evaluation of Community-Based Programs." *Evaluation Review*, 1998, 22 (1), 76-92.
- Robert Wood Johnson Foundation. "Fighting Back: Community Initiatives to Reduce Demand for Illegal Drugs and Alcohol (Call for Proposals)." Princeton: Robert Wood Johnson Foundation, 1989.
- Saxe, L., and Fine, M. *Social Experiments: Methods for Design and Evaluation*. Thousand Oaks, Calif.: Sage, 1981.
- Saxe, L., and others. "Think Globally, Act Locally: Assessing the Impact of Community Based Substance Abuse Prevention." *Evaluation and Program Planning*, 1997, 20 (3), 357-366.
- Scriven, M. "Truth and Objectivity in Evaluation." In E. Chelmsky and others (eds.), *Evaluation for the 21st Century: A Handbook*. Thousand Oaks, Calif.: Sage, 1997.
- Shadish, W. R., Jr., Cook, T. D., and Leviton, L. C. *Foundations of Program Evaluation: Theories of Practice*. Thousand Oaks, Calif.: Sage, 1991.
- Spickard, A., Dixon, G., and Sarver, F. "Fighting Back Against America's Public Health Enemy Number One." *Bulletin of the New York Academy of Medicine*, 1994, 71, 111-135.
- Stake, R. E. "To Evaluate an Arts Program." In R. E. Stake (ed.), *Evaluating the Arts in Education: A Responsive Approach*. Columbus, Ohio: Merrill, 1975.
- Stake, R. E. "Program Evaluation, Particularly Responsive Evaluation." In W. B. Dockrell and D. Hamilton (eds.), *Rethinking Educational Research*. London: Hodder and Stoughton, 1980.
- Tobler, N. S. "Meta-Analysis of 143 Adolescent Drug Prevention Programs: Quantitative Outcome Results of Program Participants Compared to a Control or Comparison Group." *Journal of Drug Issues*, 1986, 16 (4), 537-567.
- Tobler, N. S., and Stratton, H. H. "Effectiveness of School-Based Drug Prevention Programs: A Meta-analysis of the Research." *Journal of Primary Prevention*, 1997, 18 (1), 71-128.
- Wandersman, A., and others. "Comprehensive Quality Programming: Eight Essential Strategies for Implementing Successful Prevention Programs." *Journal of Primary Prevention*, 1998, 19 (1).
- Weiss, C. H. "Introduction." In C. H. Weiss (ed.), *Using Social Research in Public Policy Making*. San Francisco: New Lexington Press, 1977.
- Weiss, C. H. "Improving the Linkage Between Social Research and Public Policy." In L. E. Lynn (ed.), *Knowledge and Policy: The Uncertain Connection*. Washington, D.C.: National Academy of Sciences, 1978.
- Wholey, J. S. *Evaluation and Effective Public Management*. Boston: Little, Brown, 1983.
- Winick, C., and Larson, M. J. "75 Community Action Programs." In J. H. Lowinson, P. Ruiz, R. B. Millman, and J. G. Langrod (eds.), *Substance Abuse: A Comprehensive Textbook*. (3rd ed.) Baltimore, Md.: Williams and Wilkins, 1997.

LEONARD SAXE is professor of psychology at the Graduate School and University Center, University of New York, and adjunct professor of the Heller School, Brandeis University. He is principal investigator for the evaluation of the Robert Wood Johnson Foundation's Fighting Back initiative.

ELIZABETH TIGHE is senior research associate at the Heller School, Brandeis University, where she is co-investigator on the evaluation of Fighting Back, a Robert Wood Johnson Foundation multisite, community-based substance abuse program focused on demand reduction.