

Valuing Evaluation

Felicia G. Cohn, Ph.D.*

Director of Medical Ethics, University of California, Irvine
College of Medicine

EVALUATION OF a social program is generally undertaken to demonstrate the value of a program. More formally, program evaluation refers to “the use of social research procedures to systematically investigate the effectiveness of social intervention programs that is adapted to their political and organizational environments and designed to inform social action in ways that improve social conditions” (Rossi, Freeman, Lipsey, 1999: 2). Frequently, much is made of the practical and policy implications of such evaluation results, while less attention is paid to the endeavor itself. Yet, the quest to demonstrate value is in and of itself a matter of value, i.e., something of importance.

Evaluation, then, is fundamentally an ethical enterprise: an effort to distinguish right from wrong, good from bad, and degrees of goodness or badness. Ethics is a discipline fundamentally concerned with questions of “should:” “What should I do?” “What should be done?” Therefore, evaluation serves to provide answers to those questions with regard to specific programs. The questions of what to evaluate and even whether to evaluate suggest three levels of ethical inquiry. First, on what values is evaluation founded? Second, what values does evaluation reveal? Third, does evaluation fulfill those values?

The questions arise most intensely in the context of specific inquiries. For example, in 1998, the U.S. Congress directed the Institute of Medicine (IOM) of the National Academies of Sciences and Engineering to exam-

ine the training needs of health professionals to respond to family violence in order to develop a social action strategy. Specifically, the charge directed that an interdisciplinary panel of experts assess training needs, existing training programs, and efforts to foster knowledge among health professionals. In essence, policymakers were requesting an evaluation of the state-of-the-art in order to determine how best to proceed. The request itself and the results of the committee’s assessment will be used to depict issues of value in evaluation.

The Ethical Foundations of Evaluation

Evaluation is not generally conducted for the sake of conducting evaluation. Evaluation is not perceived to have intrinsic value, at least not beyond the academic domain. Nor is it usually undertaken simply for descriptive purposes. Certainly, evaluation describes program performance, but this description contributes to the main goal of evaluation: determining effectiveness or success measured against some set of standards. The very fact that evaluation is purposive indicates that the value of evaluation is largely contingent on derived outcomes. The findings can be used, for example, to determine if a program is worthwhile or ineffective, to quantify how effective a program is, to identify aspects of programs in need of enhancement or change, and/or to describe unexpected outcomes.

That the act of evaluating has occurred is probably of little interest without the findings.

And even the results may be of little interest without some application, such as developing policy or managing a program in order “to inform social action in ways that improve social conditions” (Rossi, Freeman, Lipsey, 1999: 2). Thus, evaluation is largely a teleological enterprise, that is, whether it is good is determined by its ends. So, evaluation is good if it is likely to produce good (See e.g., Purtilo, 1993).

The *telos* or “end” of program evaluation is varied. History demonstrates that human beings throughout time have endeavored to describe, understand, change, and improve the conditions of our existence, whether these activities are called evaluation or not. Further, with efforts to change society has come a desire to determine the impact of these efforts. Programs are usually designed to raise awareness of a social problem, address specific aspects of a social problem, or to resolve a problem. Consequently, program evaluation is used to investigate the effectiveness of a particular program in achieving the goal it was established to achieve.

In evaluating, we examine the value of a program, a determination rooted in not just whether a program is implemented as planned, but whether that program works. A determination of whether the program works depends on whether the effects of that program coincide with other things we believe important, which are described as values. These values may reflect the need for the program, the program design, the services the program provides, the cost-benefit ratio, and/or the program’s impact.

*Felicia Cohn, Ph.D. served as the Study Director for the Committee on the Training Needs of Health Professionals to Respond to Family Violence of the Institute of Medicine, National Academies of Sciences and Engineering, in Washington, DC. The views expressed in this paper reflect her interpretation of the committee’s process and findings. Neither the National Academies nor members of the committee are responsible for the opinions expressed.

When Congress mandated a study on health professionals and family violence, the legislation did not provide an explanation for the desire to know. However, certain assumptions appear reasonable based on the context of the request. Family violence can be described as a growing national pandemic (Tjaden and Thoennes, 2000). Efforts to address family violence have largely fallen within the criminal justice and social services sectors, but have been limited and have demonstrated only moderate success. The nature of health professional work uniquely situates care providers to identify and assist victims. For some victims, health professionals may be the only or a rare point of human contact outside their abusive or neglectful environments. Recognizing this, every state has legislated a requirement placed on an array of health care professionals to report child abuse and neglect and elder maltreatment. Four states similarly require that intimate partner violence be reported. In addition, a few states mandate education about some type of family violence in health professional training.

Given this background, the IOM committee presumed that Congress was working from the *ethical* premises that health professionals can improve the social condition of victims and should bear (some) responsibility for the problem of family violence. Based on these premises and in light of its charge, the committee believed the following questions to be at the heart of its task:

- 1) What is known about the response of health professionals to family violence?, and
- 2) How should health professionals respond to family violence?

The *telos* of this inquiry appeared to be a description and assessment of program designs, services provided, and impact.

The Values Evaluation Reveals

Values arise from evaluation in at least two ways. A request for evaluation in and of itself suggests that the area of inquiry is important, valuable. The evaluation protocol, how it is conducted and its findings, also communicates values. In requesting the study, Congress clearly indicated the importance of family violence and the role of health professionals in addressing it in our society. What the IOM committee found, as a result of evaluation, however, sends a message that is less clear.

The National Academies have built a reputation for providing comprehensive analysis of the existing evidence base on a topic of inquiry. Within those established methodological parameters, the IOM committee responded to its charge by consulting the existing literature, including policy and guidelines on the subject; unpublished curricula; and representatives from health professional education programs, advocacy groups, policy-makers, criminal justice and social services, researchers, scholars, and funders.

The committee's assessment of evaluation data demonstrated substantial interest in the problem of family violence within society and among health care professionals, for it emphasized societal values on understanding the extent of the problem, providing services to benefit victims, and preventing victims from enduring further harm.

However, the committee's findings, based on the review of written documents and consultation with experts and interested parties, also revealed a severe limit on the evidence base necessary to develop the guidance requested by Congress. In particular, a paucity of credible evaluation data existed to support the existing health professional training programs, to improve the existing programs, or to develop new programs. Among the Committee's findings (Cohn, Salmon and Stobo, 2001):

- Family violence is understood to be widespread across the United States and to have significant health consequences, but its full effect on society and the health care system has not been adequately studied or documented.
- Numerous studies document the incidence, magnitude, characteristics, and implications of the problem, but variation in definitions, data sources, and methods has rendered unclear findings that cannot be compared.
- Several training programs do exist, but have not been adequately studied with regard to their impact on health professional knowledge and practices or the effect on the health outcomes of victims.
- Studies of the impact of mandatory reporting requirements suggest mixed results and the need for further study.
- State-mandated education requirements do not appear to have been studied.

- Funding for research, educational development, and evaluation of health professional training programs is fragmented and inconsistently available.

So, while society recognizes family violence as an important issue for health professionals as well as criminal justice and social services—a recognition of importance reinforced by the Congressional mandate—that recognition is not reflected in research, program development, evaluation of laws, or funding commitments. The methods used to evaluate family violence generally and family violence education for health professionals specifically are described as generally lacking rigor and the limited data available send a conflicting message about the priority of this social ill. If one replaces “family violence” with another issue, such as rehabilitation programs for offenders, it is readily observable that, for the most part, the IOM committee's findings are just as applicable. The same may hold for the evaluation of other social ills and may even reflect the state-of-the-art of social science research in general.

The Value of Evaluation

Rigorous and responsible evaluation clearly has value. It has become an integral tool for decision-making, especially in government and business, and has a history of demonstrated impact. However, that evaluation can be valuable does not mean that it will be valuable in particular situations or that it will answer particular questions. *Is* does not imply *ought*; that we can evaluate does not mean that we should, as evaluation is not necessarily in all circumstances good. (See Moore, 1903, for an explanation of the naturalistic fallacy.) In fact, the unreflective use of evaluation may actually undermine its value. For evaluation to be valuable, an accounting of its limitations is necessary. The following suggest the ethical limits of evaluation.

The Need for Evaluation Suggests but Does Not Define Value

Congress premised its legislative mandate for a study on family violence training for health professionals on the importance of the issue and the expected role of health professionals in it. The committee appointed to study the issue, comprised largely of researchers and practitioners in family violence and health professional education, concurred with the premise, but believed that the public consum-

ers of the final report would require an explanation and health professional educators would need justification for including family violence in their curricula. In the committee's judgment, society not only needed to recognize the magnitude of the problem (each year about 25 percent of Americans are affected by family violence), but also needed to identify family violence as a health care issue. Family violence has traditionally been seen as a private matter among family members, which, in extreme circumstances, may become a concern of law enforcement and social service officials. Given those perceptions, the committee recognized that deans of medical and nursing schools would need to justify the inclusion of family violence in their curricula. This would be particularly true as educational time constraints mean prioritizing among demands that a number of social and other issues be incorporated into the curricula. The committee sought not only to identify health professional training needs, but also to highlight the urgency and importance of the issue of family violence itself.

Evaluation, or Lack Thereof, May Reveal Priorities

While Congress backed its value claim by authorizing funds from the Centers for Disease Control and Prevention for the IOM study, the value of evaluation is not always so explicit. Specifically, funding is not always available even for programs generally thought to be highly valued. Both the program to be evaluated and evaluation itself must be of sufficient merit to justify the expense. The committee found a number of funding sources providing monies to create training programs, but few specified that funds be appropriated to evaluate those programs, even when the funder required that evaluation be done. Evaluation is generally expensive, so even thoughtful methods and visionary scope may not be enough to produce helpful evaluation, or any evaluation at all, if funds are provided to do it. This seems to suggest that creating programs to address family violence education for health professionals was more important than determining the impact of those programs (a situation similarly found throughout the field of criminal justice and other social problems).

While the funding organizations claimed that the success of the programs they supported was at least as important as the cre-

ation of the programs, actual funding sent a different message. The message may simply be that real economic constraints mean developing programs first and worrying about evaluation later. However, another possible explanation is that evaluation could provide information we would rather not have, e.g., that an expensive program is not successful or not cost-effective.

Methodological Limitations of Evaluation Can Affect its Value

Well-designed evaluation is likely to demonstrate some of the impacts of a program, provided program objectives are explicit and desired outcomes are delineated. However, conclusions about impact are likely to rest on correlation rather than evidence of causation. The IOM committee did find studies demonstrating increased knowledge about family violence among health professional trainees who participated in curricula with family violence components. This suggests but does not prove a causal relationship; that is, A does not necessarily cause B. So, too, evaluation of crime prevention programs may result in some programs being labeled "successful," while others are no more than "promising" (Sherman, et al, 1998). Evaluation can offer reasonable explanations, but affirming some values will continue to require a leap of faith.

The Results of Evaluation Can Create Misleading Value Claims

A critical eye is necessary to detect poor or biased design, inappropriate interpretation, or pure propaganda. For example, much of the "research" on family violence the committee uncovered was sponsored or conducted by advocacy organizations. This, in itself, does not negate the findings. But evaluation studies tied to organizations with vested interests in supporting particular values may result in studies that support those values. Further, in reviewing training program evaluation, the committee found that findings of success often turned on self-reports from individuals who had received the training. These subjective findings really suggested only that the survey respondents remembered and/or liked the course. No objective evaluation indicated whether those who received the training either retained or used the information; that is, whether or not the training had positive impact.

Even "Good" Evaluation May Not Tell You What You Need to Know to Improve the Social Condition.

The scope of evaluation is generally limited. Successful implementation does not mean a successful program and short-term results may not mean long-term results or the kind of impact most desired. The committee's literature review suggested that curricula architects had accomplished a great deal, but a closer look suggested the only real success was in getting a program into a particular curriculum or maintaining the educational component over time. These are certainly important, for if a program cannot be implemented or maintained, there will be no outcomes to evaluate. But the mere existence of a program does not mean it is working or working well. Similarly, findings of increased knowledge about family violence among health professional trainees do not mean that the education had an impact on the practice patterns of those trainees or on the resulting health status of victims. In addition, the context in which a program is evaluated is important, as programs of demonstrated effectiveness do not exist in a vacuum. The environments in which they are implemented bear on their success and may limit the generalizability of the results, so that a program that is very successful in a particular setting may be an utter failure in another. Thus, evaluation may not always be able to serve a desired or wanted goal of improving social conditions.

Good Intentions, Experience, and Consensus Opinion Can Be as Powerful as Good Evaluation

Good evaluation may not always be possible. Designing rigorous studies to assess the impact of family violence education on the health outcomes of victims may simply not be possible or may be cost- or resource-prohibitive. The same, of course, can be found in all social science research, including that related to correctional issues. The IOM committee, committed to assessing the existing science, was consistently frustrated in its desire to develop recommendations based on the collective knowledge and experience of its members, as well as existing consensus within the field.

Evaluation is not the only tool for identifying value. An examination of existing curricula, policy, consensus statements, and expert opinion indicated substantial overlap about the

content of family violence curricula for health professionals. The committee, instead of recommending specific content areas, could only recommend these areas as starting points for evaluation. While evaluation will be necessary to confirm these content areas, the other sources powerfully assert their value.

Conclusion—What Works

Despite its limitations, evaluation is invaluable in distinguishing worthwhile programs from the less worthwhile and unworthwhile. However, evaluation is perhaps just as valuable in signifying value. Ethical consideration suggests that for evaluation to work to its full potential, at least two levels of reflection are necessary. First, evaluation should be specific and well-formulated for development and implementation. This involves determining whether the evaluation can describe, identify, and assess that which needs to be described, identified, and assessed, and crafting methodologically rigorous tools for assessment. Misusing evaluation in an attempt to demonstrate that which cannot be demonstrated will only undermine its value. Second, the reasons for undertaking evaluation at all should also be a matter of careful thought. Because evaluation is a contingent good and

not of value in and of itself, it underscores the value of that which is being evaluated. Assuming a need to evaluate when one may not exist may also result in devaluing evaluation. What is needed are efforts to determine what works in the context of decisions about what is most important to have work.

Undeniably, all programs reflect values of one kind or another. The researchers who assess these programs must also contend with their personal values and the influence these values have on study designs and outcomes interpretations. In the final analysis, the “goodness” or “badness” of a program is reflective of decision- and policy-makers’ own value systems.

Evaluation, nonetheless, is a necessary tool to determine program efficacy and should be considered an integral component of program design and implementation. Evaluation outcomes, whether positive or negative, moreover, ostensibly become—or should become—the basis for determining a program’s future: to continue, modify, or abandon it. This, of course, is an administrative decision, which is influenced not only by a study’s outcomes and value systems, but also by superordinate requirements, politics, resource availability, needs, demands, and stakeholder interests, including the latter’s values.

References

- Cohn, F., M.E. Salmon, J.D. Stobo, eds., *Confronting Chronic Neglect: The Education and Training of Health Professionals on Family Violence*, Washington, DC: National Academy Press, 2001.
- Moore, G.E., *Principia Ethica*, Cambridge: Cambridge University Press, 1959.
- Purtilo, R., *Ethical Dimensions in the Health Professions*, Philadelphia, PA: W.B. Saunders Company, 1993.
- Sherman, L.W., D.C. Gottfredson, D.L. MacKenzie, J. Eck, P. Reuter, and S.D. Bushway, “Preventing Crime: What Works, What Doesn’t, What’s Promising,” Washington, DC: U.S. Department of Justice, 1998.
- Tjaden, P. and N. Thoennes, Report No. NCJ 181867, Atlanta, GA: National Institute of Justice, Centers for Disease Control and Prevention, 2000.