

## **An Approach to Evaluation at The California Wellness Foundation**

By Tom David  
Executive Vice President

Since its earliest days, the Foundation has been committed to evaluation of its grantmaking. That commitment is reflected in four of our core values: Accountability, Stewardship, Excellence and Learning. We see ourselves as a learning organization, and we believe that evaluation is an important aspect of the process of continuous learning and program improvement.

Our approach to evaluation has evolved over time. Initially, we required each grant to have a formal evaluation component, often funded at a level equal to that of the project itself. We also matched community-based organizations with evaluators to help them meet our requirements. As our five strategic initiatives came on line, we made substantial (typically, multimillion-dollar) grants for external, objective evaluation of each of those grantmaking programs. Our first evaluations of that sort employed a “scientific” approach, with the primary goal of contributing to knowledge.

Subsequent initiative evaluations have been shaped by a somewhat different set of goals. Our approach evolved from an attempt to “prove” to one that sought to “improve.” The Foundation has made it clear that although we are still concerned about tracking outcomes, our first priority has been to provide continuous feedback to our grantees to help them enhance program effectiveness. We have also acknowledged the importance of building the capacity of grantees to conduct their own data-gathering and evaluation activities as a key component of the ultimate sustainability of their work.

More recently, we have begun to conduct retrospective “cluster” evaluations of groups of grants made for a similar purpose. Some of those evaluations have been commissioned from investigators outside the Foundation, but we are increasingly taking on those analyses ourselves as an integral part of our grantmaking cycle. We have made it an institutional priority to dedicate time and energy to postgrant analysis and synthesis of lessons learned in order to inform our future work.

Lest this leave the reader with the impression that we have “solved” the dilemma of how best to evaluate our work and that of our grantees, let me pause here for a disclaimer. We are still very much a “work in progress” on this topic as on others. The purpose of this introductory piece is to share some of our thinking to date and to provide some context for the evaluation products that will be posted in this section of our website. All of this information is shared in the spirit of mutual learning. We welcome your comments on the specific reports as well as on this contextual piece.

**Why engage in evaluation?** After all, even if fairly modest sums are allocated for that purpose, those are still dollars that could otherwise be used for grants to agencies to deliver health services. In her introductory chapter to the Council on Foundations’ 1993 volume on *Evaluation for Foundations*, Barbara Kehrler outlines several reasons why foundations might invest in evaluation, among them: accountability; program improvement; assessment of program impact; dissemination of innovative programs; and increasing the state of knowledge.

Accountability is frequently framed in terms of the grantees' activities; i.e., did they deliver the goods as promised? One of the most important principles of The California Wellness Foundation is a commitment to good stewardship of the Foundation's resources. In other words, our impetus for evaluation is to examine whether we are making the most effective use possible of our grant dollars. That constitutes our "bottom line" for accountability.

Some contend that foundations that have resulted from health care conversions should be held to a higher standard of accountability than foundations created from other sources, since their assets are in essence derived from "public shareholders." While we would not necessarily agree with that argument, this is an organization that takes very seriously its community responsibility as a charitable institution. One could argue that if our mission is to improve the health of Californians, then we are ultimately accountable to the communities served by our grantees. If it is conducted appropriately, evaluation can be a useful tool to help fulfill that commitment.

In an article some years ago in *Foundation News*, Humphrey Doermann (then president of the Bush Foundation) suggested four potential levels of evaluation for foundations:

- The overall functioning and impact of the foundation itself. This is perhaps the rarest form of foundation-sponsored evaluation, yet it is why we conduct periodic surveys of our grantees and applicants. It is also one reason why we have instituted the *Reflections* publication series, to communicate some of our lessons learned with the field. Those who have commented on that series have frequently noted how unusual it is for a foundation to openly share such a self-assessment with a broader audience.
- Impact on a particular field of endeavor. Somewhat further down the road, one could imagine us commissioning an assessment of the cumulative impact of our grantmaking in our various priority areas on their respective fields of interest across California.
- Accomplishments of a particular grants program. This is why we have evaluation grantees for each of our initiatives. It is also one of the reasons why we have had Advisory Committees for most of our initiatives, to give us independent feedback on their implementation and impact. We have also commissioned retrospective analyses of groups of grants within our Special Projects program, and plan to do more of these in the future for our other priority areas.
- Effectiveness of individual grants. The primary vehicle here is typically monitoring by the program officers, with the outcomes analyzed in a closeout report, a summary of which is reported to the Board. If the magnitude of a grant warrants it, a special evaluation component might be built in or commissioned by an outside investigator.

### **What have we done so far?**

We have invested substantial sums in independent evaluations of each of our five strategic initiatives. In each case, the evaluation grantees were selected by competitive RFPs, and they represented a variety of organizations from universities to private, for-profit corporations. They were charged with providing both the Foundation and our grantees with regular feedback on

program development and effectiveness, summarized in an annual progress report. Each evaluation grantee was also expected to deliver a final report at the conclusion of the initiative to address not only the outcomes achieved by individual grantees but also the overall impact of the initiative itself.

Parenthetically, as a result of our experience with initiative evaluations, we are now clearer on what we expect in the way of a final report. It should be no more than 50-60 pages and written in “plain English” in a user-friendly graphic format. Such a document could then potentially be disseminated to a variety of audiences and stakeholders outside the Foundation itself.

As for grants made outside of our initiatives, we no longer require that each have an independent evaluation (as we did in earlier days), but we do expect them to have explicit objectives and a way of measuring progress toward their attainment. On a couple of occasions, we have also commissioned an outside investigator to conduct a post-hoc “cluster analysis” of a group of grants made for a similar purpose to get a better sense of their overall impact. Such evaluations are low-cost and can be completed in a relatively short period of time. One of the most recent examples of such an analysis is Jennie Schacht’s report on grants made to regional clinic associations and consortia, which is posted in this section of the website.

“Cluster” evaluations can’t provide “definitive” data, but they can offer reasoned judgments about the effectiveness of our funding. They can also begin to capture more of the “story” of what happens as a result of our grants. Which raises a third question:

**What questions would we like evaluation to answer?** A logical first step in evaluation is to achieve a clear description of the program as it was actually implemented “on the ground.” What happened at each site as a result of our funding? That question argues for an in-depth, qualitative assessment that captures not only the basic numbers of clients served, but more importantly, the “texture” of the program. Such an evaluation requires skilled observers and interviewers who are attuned to the cultural context of the organizations and the participants. That kind of analysis is also most likely to surface the participants’ true feelings about a program and how it might be improved in a way that typical questionnaires might not. It can also help document changes in the program itself over time, in response to corrective feedback.

As we are moving to grant programs that focus on the sustainability of grantee organizations, another important focus for inquiry is the measurement of institutional capacity. Have we strengthened the participating organizations? Too many evaluations focus on the “theory” of the intervention, ignoring basic dimensions of organizational functioning that can be absolutely key to the quality of the programs, and thus the ultimate success or failure of the project. Although there are not many examples in our field, we have begun to look across other sectors of philanthropy (e.g., youth development, the arts) for examples of grant programs whose primary purpose has been to build organizational capacity. There is much to be learned from their experience to inform our next steps.

Much of the recent interest in evaluation among foundations has been in response to critiques raised both by board members and foundation observers that we have paid insufficient attention to the

measurement of the outcomes of our grants. What was actually achieved? To be fair, the state of the art of assessment in health and human services, whether funded by government or private sources, has concentrated on process variables such as the number of patients treated, the number of classes presented, and short-term results such as scores on “post-tests.” Given the nature of much of our work—primary and secondary prevention—many of the ultimate outcomes we seek are not readily measurable. Is it sufficient to settle for intermediate measures? Which raises an additional question:

**What is a realistic “standard of evidence”?** In response to this question, a methodologist might begin by sketching out a simple continuum of data sources. At one end, one might put uncorroborated self-reports from grantees, coupled with semi-random Program Officer observations. At the other end, one would find true experimental designs, featuring double-blind random assignment of “subjects” to “treatments.” In between those two poles would fall an array of intermediate alternatives, ranging from qualitative observations by evaluation professionals from outside the organization to quasi-experimental time-series designs utilizing comparison groups.

When I was studying evaluation as a graduate student, there was no question in my instructor’s mind as to which method was most desirable. It was experimental rigor or nothing! But the field of evaluation has evolved over the past 25 years, and now even the experts disagree on the answer to this question. The key word here is “realistic.” It depends on what outcomes you are hoping to achieve.

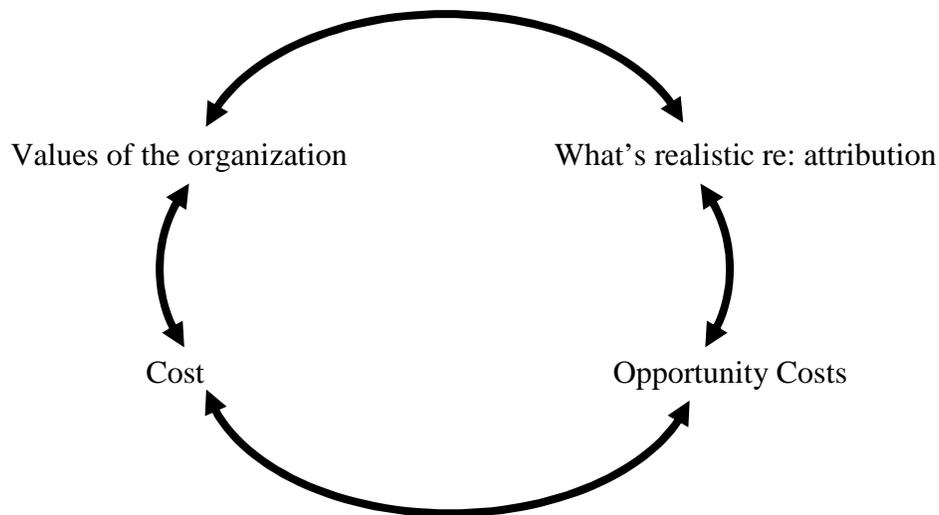
Some health outcomes are relatively easy to measure in a fairly short time frame—for example, whether a prenatal outreach program has been able to reduce the number of women delivering babies with inadequate prenatal care in a particular community. But long-term changes in individual behavior, or some of the community-wide outcomes to which we aspire as an organization, can be virtually impossible to track with precision, let alone during the limited life of a funding initiative.

Moreover, many of the most methodologically ambitious attempts to evaluate long-term program impact have yielded disappointing results, feeding the perception in some quarters that “nothing works.” Yet if we step back a bit from our work, it stands to reason that it’s rather unrealistic to expect time-limited programs to engender long-term change, particularly in communities with few other support systems in place. That is why we and others have invested in longer term, multifaceted funding initiatives. But it only makes the challenge of evaluation that much more complicated.

Even with a relatively sophisticated evaluation design in place, there remains the challenge of attribution. How do we know that the results observed are due to the program we’ve funded? As we’ve noted before, most of our grant programs are being implemented in “high noise” settings, where multiple interventions are simultaneously taking place. Even if we were able to employ methodologies such as random assignment and control groups, there’s no guarantee that we would be able to unequivocally attribute observed outcomes to our funding.

I would recommend reserving such efforts for only the most “mature” programs, where data of this sort might be critical to assessing their potential for widespread replication. As an example, we cofunded an ambitious random assignment field trial in San Diego of a previously tested model of home visiting for families with newborns. One of our partners was the California State Department of Social Services, which was considering this model for significant statewide funding. It made sense to take the time and expense to underwrite a sophisticated experimental design in that case, since the model, although successful in other settings, had not yet demonstrated its effectiveness with underserved populations in California.

Going back to the reasons for engaging in evaluation for a moment, there’s only one – advancing knowledge—that demands the “gold standard” type of experimental design described above as the acceptable standard of evidence. Rarely in the worlds of policy and practice are such “textbook” standards decisive—for better or for worse. Judgments tend to be made on other forms of information, whether they are quantifiable, intermediate measures of success, other forms of documentation or even well-told anecdotes. Does that mean we should abandon rigor for pragmatism? To me, as in other aspects of philanthropic decisionmaking, it’s a matter of institutional values and tradeoffs, as expressed in this schematic:



As an organization, we tend to place a higher value on strategically meeting the health needs of underserved communities than we do on contributing to knowledge. Those endeavors are not mutually exclusive, of course, but their prioritization is instrumental in shaping the design of our grant programs and what we ultimately fund. It also determines the potential audiences for our work. Whom do we hope to influence, and what forms of data will be most persuasive?

As a result of our recent strategic planning process, we have charted a course for the future that moves us from our original “paradigm” that focused on idea/theory testing to one that prioritizes sustainability of key organizations around the state. That suggests an approach to evaluation that emphasizes an in-depth, qualitative description of the work, ongoing feedback for program

improvement, and outcome measures related to organizational capacity as intermediate steps to improving the quality of health-related programs and policies.

Those considerations, in turn, have implications for a realistic approach to measurement that recognizes the challenge of attributing observed changes to our funding and for the cost of evaluation activities. The implicit balancing point is the ultimate utility of the findings. How will the final product actually help to improve practice, either by our grantees or the Foundation? Those factors then need to be weighed against the opportunity costs—the other potential uses of dollars that we might devote to evaluation.

There is no clear “industry standard” among foundations regarding the dollars allocated to evaluation. Some large foundations such as Robert Wood Johnson and W.K. Kellogg have substantial in-house evaluation departments. In some cases, those staff conduct evaluations themselves, but they typically serve as advisors to program staff and brokers in the funding of external evaluations. Costs can be significant, although the impact of much of that work is rarely felt outside the organizations themselves. An informal, strictly nonscientific survey of a number of other funders that have chosen to invest in evaluation indicates that it’s more typical to devote between five and eight percent of their grantmaking budget to evaluation.

Although it’s often overlooked, it is also important to consider the costs incurred by our grantees themselves when they are required to participate in evaluation. They must devote considerable time and resources to trainings, data collection, reporting and on-site meetings with the evaluators. And that is all in addition to other meetings with technical assistance providers, public education grantees, and Program Officers that they must engage in as part of an initiative. Those costs are not insubstantial, and should also be considered in determining an institutional approach to evaluation.

Given all of the above, the Board of The California Wellness Foundation adopted the following principles for evaluation in December 1999:

### **Principles for TCWF’s evaluation activities**

- Central to our conception of stewardship is a commitment to hold ourselves accountable to high standards regarding the quality of our work and the effectiveness of our grantmaking. We expect our grantees to hold themselves to a similarly high standard of performance.
- We recognize the complexity of the task of evaluating the kinds of programs we fund as well as the tradeoffs involved in determining an appropriate funding level for those activities. Consequently, our expectations regarding evaluation are tempered by that understanding.
- Our commitment to fund in traditionally underserved communities brings with it some special evaluation challenges, including uneven institutional capacity for evaluation, a lack of patience with “being studied” given past negative experiences, and political sensitivities regarding premature judgments of “failure” that could have devastating consequences for local residents.
- A central theme of our grantmaking is to enhance the sustainability of front-line providers of health services as well as organizations working to influence public policy and those whose

mission is to nurture leadership. We are willing to take informed risks to fund organizations that are struggling against the odds to provide essential services to their communities. Evaluation can be a useful tool to assist them in those efforts.

- We support an approach to evaluation that emphasizes building skills among grantees for self-assessment and continuous feedback for program improvement. An in-depth, qualitative description of program implementation is an important complementary activity.
- It is also important to do the best we can to measure outcomes of our work. We recognize that funders have often required grantees to promise too much in the way of outcomes for the dollars received, which serves no one well. Instead, we hope to help our grantees set realistic goals, just as we understand the limits of evaluation to clearly attribute observed changes in the environments in which our grantees operate to our funding alone.
- We are committed to candid communication about the lessons learned from those activities, both internally and to key external audiences. We see evaluation as a means to the end of creating a true learning community, both among our grantees and among ourselves and our colleagues, where the institutional culture supports a spirit of genuine inquiry rather than the fear of judgment.
- Recognizing the limits of our grant dollars in the context of the multiple health needs of Californians, we choose to balance our investment in evaluation in terms of the ultimate utility it will provide vs. the costs incurred, including the time and energy required of our grantees. As a general guideline, we plan to invest no more than five percent of our grantmaking budgets to that end.