

The Robert Wood Johnson Foundation's Approach to Evaluation



JAMES R. KNICKMAN AND KELLY A. HUNT

Robert Wood Johnson
Foundation

Chapter Nine,
excerpted from *The Robert
Wood Johnson Foundation
Anthology*:

**To Improve Health
and Health Care
Volume XI**



Edited by
Stephen L. Isaacs and
David C. Colby
Published 10/2007

Editors' Introduction

Foundations, as Michael Porter and Mark Kramer observed in an influential article published in 1999, are often ambivalent about evaluation. “The overall failure to evaluate the results of foundation grants is the most telling danger sign of all. Almost no money is set aside for program evaluation...[yet] without evaluation, a foundation will never know whether or not it has been successful.”¹ Evaluation has become more important in the field of philanthropy since then, as foundation boards and staff members increasingly demand evidence of the impact of their grant-making and as policy-makers and the public insist on more accountability.

The Robert Wood Johnson Foundation has placed a high priority on program evaluation since its inception as a national philanthropy in 1972. It has developed a four-tiered system of evaluation that ranges from the evaluation of individual grants and clusters of grants to the qualitative assessments found in *The Robert Wood Johnson Foundation Anthology* series.

In this chapter, two architects of the Foundation's evaluation system describe that system and examine its strengths and weaknesses. As the Foundation's vice-president for research and evaluation between 1992 and 2006, James R. Knickman was the person primarily responsible for the Foundation's approach to evaluation. Kelly A. Hunt, who was at the Foundation between 2000 and 2006, was a research officer and in charge of the Foundation's “Scorecard.” Knickman is currently the president and chief executive officer of the New York State Health Foundation, and Hunt is a senior program director there.

This is the first comprehensive public presentation of the way that the Robert Wood Johnson Foundation assesses its programs and their effectiveness. As such, it may resonate with other foundations and nonprofit organizations seeking to have a better understanding of the impact of their activities.

1. Porter ME and Kramer MR. “Philanthropy's New Agenda: Creating Value.” *Harvard Business Review*, 121–130, Nov.–Dec. 1999.

In philanthropy, as in business, the bottom line counts. But each sector measures the bottom line differently. Businesses can look to revenue, income, sales, earnings per share, and other quantitative indicators to measure performance. Foundations often work with less clear-cut indicators of impact. Sometimes they find it difficult to quantify the objective they wish to achieve. At other times it is difficult to know whether social change is due to the actions of a foundation or to other forces.

Over the past 20 years, the imperative to evaluate outcomes and assess effectiveness has become progressively important in philanthropy. Government leaders and others who watch philanthropy increasingly push for evidence that foundations use their resources wisely and actually contribute to addressing social problems. Boards of trustees at foundations have also become increasingly demanding when it comes to accountability and impact.

This imperative for evaluation is most characteristic among foundations that attempt to effect social change in active, coordinated ways. If a foundation chooses to focus on direct charity—funding services at homeless shelters, say, or supporting free medical care for the uninsured—the links between social contributions and the foundation’s resources are generally considered self-evident. The impact of such contributions can be measured by the number of people served or by improvement in their health. When a foundation focuses on making grants that address a social problem, such as efforts to end homelessness or to bring about universal health insurance coverage, then it becomes more difficult to demonstrate a causal relationship between a foundation’s investment and a change in the social environment.

The Robert Wood Johnson Foundation is generally regarded as a foundation that takes evaluation and performance assessment seriously. It has a staff of 12 professionals plus support staff members focused on evaluation, and it has been conducting evaluations of grants since its earliest days, in the 1970s.

Historical Roots

The idea of evaluating the effectiveness of grant programs came naturally to the early leaders of the Foundation. The board of directors consisted largely of recently retired Johnson & Johnson executives and others who had been colleagues of the founder. Accustomed to drug trials in which the effectiveness of a new pharmaceutical product is rigorously tested before being placed on the market, the board looked for comparable approaches to judging the impact of Foundation programs. Many of the Foundation’s earliest investments were multisite demonstrations of new approaches to improving access to health care. Multiple states or communities or providers would be funded to try a specified strategy for, say, improving access. This type of grantmaking warranted investments in program evaluations to determine whether or not the new approach in fact had beneficial impacts. Given the interest of the Foundation in identifying ideas that could, if successful, serve as models, evaluation emerged as a central feature of its approach to grantmaking over the years.

Additionally, the 1970s, when the field of public policy analysis was emerging, were years of ferment in the area of social science research. It was natural to borrow the evaluation methods

being used by the federal government to test new approaches to delivering services to low-income individuals (“social experiments”). In the 1970s, for example, the federal government financed a range of social experiments that tested new ideas in welfare reform, national health insurance models, and workforce training.

The emphasis on formal evaluations of large-scale programs continued through the 1980s, and it continues today for programs that test new ideas for improving health care or promoting the public’s health. During the 1980s and 1990s, however, the Foundation’s grantmaking approaches became more diversified as it funded research to understand health challenges such as the lack of insurance coverage; communications efforts to increase public awareness about a range of health problems and potential solutions to them; and wide-ranging programs to address health problems such as smoking. The board also applied pressure to broaden the approaches used by the Foundation to assess impact. Board members asked two types of questions: What are we learning from all of these grants? and How can we be sure that our grants are really making a difference in improving health or health care?

As a result, in the 1990s evaluation at the Foundation began to change in important ways. Instead of just asking whether a specific grant program was effective, the Foundation began looking at groups of grants that were meant to affect a specific health problem and assessing whether the grants as a group were effective in addressing it. It developed a family of evaluation tools focused on different aspects of impact. The Foundation currently uses a four-tiered approach to evaluation.

- The first tier attempts to understand the effectiveness of specific programs. Following its well-established pattern, the Foundation hires outside institutions to evaluate the results of its major grant initiatives.
- The second tier attempts to understand the impact of clusters of programs that focus on a particular goal or set of goals. In 2003 the Foundation developed an impact framework that sets short-, long-, and medium-range targets in specific program areas (such as health insurance coverage, childhood obesity, and public health).¹ Concurrently, it developed performance indicators to measure progress toward those targets.
- The third tier examines how the Foundation as an organization is doing, using a Scorecard that is presented to the board annually. The Scorecard incorporates the impact framework’s performance indicators and commissions surveys to find out what grantees think of the Foundation, what experts think the Foundation’s impact on health is, and what the staff considers to be the Foundation’s strong and weak points.
- The fourth tier assesses the work of the Foundation in a less formal way and presents the results to a broad public. This collaborative program-evaluation-communications effort uses two vehicles: (1) grant results reports, in which the Foundation commissions writers to investigate specific grants and grant programs and write reports on them that appear on the Foundation’s Web site, and (2) *To Improve Health and Health Care: The Robert Wood Johnson Foundation Anthology*, the book series published annually by Jossey-Bass. Both of these vehicles focus on disseminating what the Foundation has learned from various aspects of its grantmaking.

Tier 1: Measuring the Impact of Specific Programs

If a state government requires schools to report to parents on their child's body mass index, will it result in lower rates of childhood obesity over time? If frail elders are given an option to hire their adult children or neighbors as caregivers instead of requiring the elders to use formal long-term-care workers, will outcomes improve and costs decrease? If a hospital provides readily accessible translation services for non-English-speaking patients, will they become healthier?

These are the types of questions that can be answered through carefully designed program evaluations. For many years the Foundation has funded independent researchers at universities and other organizations to conduct studies that assess whether the interventions it has funded improve outcomes. It is not coincidental that the Foundation relies on outside evaluators. Having an external, independent team measuring outcomes keeps the process honest. It is too easy for a foundation's program staff and its grantees to become overinvested in a program and reluctant to admit that they have not achieved the desired result. Moreover, when an intervention is successful, the judgment of an independent team increases the credibility of evaluation findings.

Program evaluations also look at implementation issues to learn why outcomes are reached or not and to document how a grantee goes about working on an initiative. Implementation analyses sometimes can help grantees midstream, and most important, they create a roadmap of do's and don'ts that can help in the replication of successful programs.

Program evaluations funded by the Robert Wood Johnson Foundation generally use social science and epidemiological research design concepts. At the heart of this tier of evaluation is comparison: whenever possible, the evaluation team compares outcomes at sites supported by the Foundation with those at similar sites not supported by the Foundation. In the ideal program evaluation, either random assignment or some other mechanism is used to ensure that the comparison sites are as similar as possible to the Foundation-supported sites.

Over the years the Foundation has learned how difficult it is to mount evaluations in the real world. The perfect evaluation design is generally elusive, baseline data may not be available, and second-best strategies for selecting comparison groups and measuring outcomes are often necessary. Even with the best of intentions, programs tend to evolve over time as priorities change, expected outcomes are revised, or grantees shift their focus. Tensions between program staff and evaluation team members can complicate the evaluation process, and the burden on grantees and program staff members to meet the needs of evaluators is frequently greater than was projected at the start. Often, the findings that emerge are inconclusive or are reported too late to influence decisions about next steps in the Foundation's grantmaking or to inform public opinion.

Thus, the practical difficulties in carrying out rigorous program evaluations are enormous. The evaluation of the *Fighting Back* program, a nearly \$100-million initiative to foster community coalitions to reduce substance abuse in high-risk neighborhoods, provides a classic case study. The program changed course in midstream; the evaluation began after the program was halfway through; the composition of the communities changed; the program office and the evaluation team could not agree on the indicators to be measured; and when the evaluation was finally completed, its findings were strongly disputed.²

Whatever the practical difficulties, evaluations do result in learning. Both the Foundation and those outside it use the results of evaluation to develop programs. SUPPORT and *Cash & Counseling* offer two examples of the practical use of evaluations.

SUPPORT—the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments—tested new approaches to providing care for terminally ill hospitalized patients; it involved patients, their families, nurses, and physicians in determining the kind of care that the patients would receive toward the end of their lives. As the demonstration project was being carried out in five hospitals, there was a sense among those involved in it that patients and their families were in fact making better choices, and that their wishes were being respected more. When the research findings were tabulated, however, clear evidence emerged that patients' wishes were not respected and that their care was no different from that of patients who did not receive the special intervention. These findings led to a fundamental national rethinking of end-of-life care and paved the way for a decade of Foundation-funded initiatives to reshape the care of terminally ill patients.

The evaluation of the Cash & Counseling program compared the satisfaction levels of people receiving long-term-care services from home health care agencies supported by Medicaid to those of another group that was given the option of using family, friends, or community members to deliver the services. The findings of the evaluation demonstrated much higher satisfaction rates among the people given the new option. Timing and reliability of care, treatment of beneficiaries by caregivers, and performance of tasks by caregivers were also substantially better for this group. The evaluation findings provided a basis for recent legislation that expands the Cash & Counseling approach from a three-state demonstration project to an option available in all 50 states.

Evaluations suggesting that no impact was associated with an initiative have led the Foundation to shift priorities. For example, when an evaluation concluded that a fellowship program to improve the understanding and teaching of health care financing was ineffective, the Foundation decided to discontinue the program.³ Similarly, a negative evaluation of a program to encourage community partnerships to lower the cost of health insurance led the Foundation to end the program and re-assess the approach.⁴

The Foundation's long experience in program evaluation has led to numerous lessons, among them:

- It is essential to design programs and evaluations at the same time so that a program can be implemented in a manner that makes an evaluation feasible.
- It is also essential to get timely baseline data if before-and-after comparisons are a feature of an evaluation.
- Evaluation findings need to emerge in a timely manner. Great insights that arrive too late to influence the next steps or federal or state policy have no impact.
- Evaluation design needs to be flexible enough to adjust to changing features of the program being evaluated. Few programs are implemented as planned, and most end up being less ambitious in scope than originally projected.
- The evaluation team and the program team need to work together. Whenever tension, a lack of respect, or an inability to work together characterizes a demonstration initiative, it is unusual for useful findings to emerge.

- Evaluations tend to be stronger when a program has clearly defined and clearly measurable expected outcomes.

Tier 2: Performance Indicators: Tracking the Impact of a Portfolio of Programs

Although program evaluations can measure the success or failure of individual programs, the Foundation's trustees and staff want to know what impact its grantmaking in a specific *field* is having. For example, when the Foundation sets an objective such as reducing tobacco use, they want to know whether smoking has gone down and, to the extent it is possible to know, what contribution the Foundation's programs have made to the decline.

Grantmaking at the Robert Wood Johnson Foundation is mainly directed at attempting to achieve tangible improvement in some aspect of the health system or health-related behavior. The staff and the trustees identify a behavior or public health, health care system, or policy issue that is of concern—such as tobacco use, childhood obesity, or the rising number of uninsured persons—agree upon a specific strategic objective (such as lowering the number of uninsured people by a certain percentage within a specified time period), and then decide on an array of grants that they expect will meet the strategic objective.

To assess whether progress is being made, staff members set down a series of performance indicators that must be reached for the strategic objective to be achieved. These performance indicators are used by the board and the senior staff to judge performance and to guide decisions about resource allocation. Objectives are modified from time to time—often in three- to five-year intervals—but sometimes a priority remains in place for 10 or more years.

The performance indicators flow from logic models that the staff develops. These specify, in simple terms, what has to happen in the short- and mid-term and then the intermediate run if the Foundation can expect to achieve a specified objective in the long run. Short-term indicators are meant to be roughly annual checks on progress that help clarify the Foundation's immediate plans and strategies. A short-term indicator for the tobacco-reduction goal might be convening a summit meeting to help frame policy priorities, followed up by a coordinated plan of action developed by grantees working in this field. Intermediate indicators are measures—often more ambitious than short-term performance indicators—that the Foundation uses to determine progress over a two- to three-year time period. As one example, the Foundation looked at the proportion of the population covered by clean indoor air laws as an intermediate indicator of progress toward its goal of reducing smoking. Long-term indicators reach beyond three years and are broader in focus and level of impact. In the area of tobacco, the Foundation tracked the prevalence of youth smoking as a long-term indicator. Table 9.1 presents a performance indicator report—with short-, medium-, and long-term indicators—that the Foundation staff and board used to track progress in tobacco control.

Table 9.1. Strategic Indicators: Tobacco

Tracking Indicators

Indicators	Control	Target Date	Baseline Status	Current Status	Target
State tobacco control funding as a percentage of the CDC minimum (tracking backslide)	M	7/08	40.2% (2000)	22.11% (as of 12/05)	27.0%
Prevalence of cigarette use among 12th graders (tracking backslide)	L	7/08	31.4%	23.2% (as of 12/05)	25.0% (as of 12/05)
“Strength of Tobacco Control” Index-monitors tobacco control capacity at state level (tracking backslide)	M	7/08	0 (Normalized/2000)	2.4 (as of 12/04)	2.4

Short-Term Indicators (4/1/05–4/1/06)	Control	Target Date	Baseline Status	Current Status	Target
Implemented fund-raising strategy with management	H	7/05	Not completed	Completed (7/05)	To complete
Completed transition plans and budgets for major programs	H	10/05	Not completed	Completed (10/05)	To complete

Intermediate Indicators (13–36 Months)	Control	Target Date	Baseline Status	Current Status	Target
Combined average state and federal tobacco excise tax	H	7/06	\$1.11 (2003)	\$1.33 (as of 9/06) Completed (10/05)	\$1.25
Proportion of the population covered by comprehensive clean indoor air laws	H	7/06	21.8% (2003)	33.1% (+.09%, 6/06) Partially completed	35.0%
Number of states that cover tobacco dependence treatment through Medicaid (tracking backslide)	L	7/06	38	41 (2003) Completed (12/05)	35 (as of 12/05)

Long-Term Indicators (+36 Months)	Control	Target Date	Baseline Status	Current Status	Target
Prevalence of youth cigarette use (high school students, grades 9–12)	L	7/07	28.5% (2001)	23.0% (+1.1%, 6/06) Completed (7/04)	23.0%
Prevalence of adult cigarette use	L	7/07	23.2% (2000)	21.1% (+0.2%, early release data as of 6/06)	18.0%
Number of states dedicating CDC-recommended amounts of MSA/tax dollars for tobacco prevention/control	L	7/08	4 (2003)	4 (as of 12/05)	10

Control: L = Low, M = Medium, H = High

Source: Report to the board of trustees of the Robert Wood Johnson Foundation, October 2006.

Each performance indicator—whether short-, intermediate, or long-term—is also assigned a level of control or influence that indicates how much effect the staff thinks the Foundation can have on reaching it. For example, it is well within the control of the Foundation to organize a meeting of experts on national health insurance, so it would be given a “high” level of control; reducing the number of uninsured by, say, 50 percent depends on factors beyond the Foundation’s control and would thus be given a “low” level of control.

Developing this performance assessment system was a major undertaking for the staff. Each internal Foundation team assigned to a specific area of grantmaking was asked to specify the concrete objectives it sought to achieve (such as reducing tobacco use by children, improving retention of nurses, or reducing the disparities among racial or ethnic groups in the care they receive for cardiac conditions). The outcome specified by the objective had to be measurable so that the Foundation’s board and the senior staff would know whether the Foundation’s grantmaking had had an impact.

The performance measures serve as signposts that help the staff and the board assess whether or not positive change is likely to occur over the long term. If short- and medium-term targets are not met, then the long-term objectives are unlikely to be met as well. The assessment of short- and medium-term objectives can lead the Foundation to alter its strategy (and the logic model guiding it), increase the level of the intervention, or perhaps reconsider the feasibility of reaching the ultimate goal.

Performance measurement is difficult. It is challenging to identify tangible targets that can be measured and that can actually change in the short or intermediate run but do not seem trivial. In managing the process, there is constant concern that focusing on measurable outcomes could lead the Foundation to address less important, though more easily measurable, problems or to adopt less risky program strategies.

The Foundation uses the performance measurement system to force more systematic thinking and to present clear choices. It also forces the staff to concentrate on common, agreed-upon goals. Although the performance measurement system provides guidance for decision-making and resource allocation, it is not followed slavishly. The board is sometimes willing to approve grant initiatives even knowing that it will prove difficult to measure the results of the Foundation’s investment. For example, it is difficult to demonstrate concrete results of the Foundation’s human capital portfolio, but the Foundation’s staff and board believe that it is important. Of course, there are critics on both sides of the Foundation’s approach to performance measurement. Some at the Foundation feel that the system drives out attention to important objectives that are difficult to quantify. Others feel that the Foundation still does not insist enough on measuring impact in some areas of its grantmaking.

**Tier 3: The Scorecard:
Assessing Organizational
Effectiveness**

The Foundation’s annual Scorecard has become an integral part of its self-assessment—a tool for senior management, the staff, and the board to assess how the Foundation is doing in a number of areas. It creates a time for formal reflection on the organization’s performance and also provides staff members and trustees with a method of identifying and addressing weak areas of organizational

performance. For more than a decade the Scorecard has represented an important tool for holding the Foundation accountable to its mission and its guiding principles.

In creating its balanced Scorecard, the Foundation adapted a concept developed by the Harvard Business School Professor Robert S. Kaplan and the businessman David P. Norton for measuring performance in the business world.⁵ Typically, in the business sector these measurements include financial, internal business, innovation and learning, and customers. The Robert Wood Johnson Foundation, whose bottom line is social change rather than profitability, translated these measurements into program impact, program development, customer (that is, grantee) service, and human/financial capital. It also incorporates, as an appendix, a review of grants management performance.

Program Impact

The first major section of the Scorecard considers whether the Foundation is meeting the goals it set for itself, by presenting the performance indicators for each portfolio and program area. Performance indicators are summarized in terms of the percentage that were completed on time, late, partly completed, or not completed at all. Some commentary on the stability of the indicators is included here, too—whether small or even major changes were made to any indicators in a particular programming area. Although some amount of change to indicators is to be expected in order to maintain the flexibility to adapt to changes in the environment, a complete lack of stability could indicate a lack of sound programming at the beginning.

Also incorporated in the program impact section are data from external audiences: grantees and outside “thought leaders”—a group that includes heads of prominent health organizations, academics, public health officials, Medicaid officials, federal policy-makers, state legislators, and the health media. The Foundation’s grantees, who have a wide range of expertise, include health researchers, practitioners, advocates, decision-makers and executives. Their judgments about the Foundation’s presence, priorities, and effectiveness are an important gauge for the Foundation. Grantees rate the Foundation’s impact in a number of ways, including:

- Impact on the field, advancement of knowledge in the field, and effect on public policy in grantees’ fields;
- Objectivity of programming and disseminated materials;
- Skill and knowledge of staff;
- Influence of programming and communications on health care leaders and policy-makers.

Thought leaders are asked some similar questions, though they are also specifically asked to rate the Foundation’s impact on particular programming areas, such as tobacco use, public health, and health insurance coverage. They are also asked to indicate their confidence in, and the usefulness of, information produced by the Foundation.

The grantees and thought leaders survey allows the Foundation to compare itself to other foundations. The Center for Effective Philanthropy conducts the grantee survey and directs similar questions to grantees of the Robert Wood Johnson Foundation and other foundations. The Center

can then let the Foundation know how it stacks up with other foundations. The thought leader survey asks respondents to rate the Robert Wood Johnson Foundation with a set of its peers, including the Henry J. Kaiser Family Foundation, the W.K. Kellogg Foundation, the Commonwealth Fund, the Pew Charitable Trusts, the California HealthCare Foundation, and the California Endowment. In sum, this section of the Scorecard provides the staff and the board with both internal and external indicators and assessments of the Foundation's impact.

Program Development

The Scorecard's program development section examines the strength of the Foundation's programming efforts, such as the soundness of its strategies and whether its interests are in line with those of its grantees, thought leaders, and the public. For example, the program development section reviews whether the public perceives health care as a major priority for the government to address, as well as specific health concerns—such as cost, quality, and the uninsured—to get a sense of whether the public's and the Foundation's priorities are aligned. The reasons for including this information in program development are twofold. First, though Robert Wood Johnson is a private foundation, it considers itself to be accountable to the public interest. Second, information about public perceptions of important health topics—childhood obesity, for example—can help guide the Foundation's programming. Moreover, if a topic is of great concern to public health professionals but does not resonate with the public or policy-makers, it is an opportunity for the Foundation to inform the public through its communications efforts.

The program development section also incorporates thought leaders' opinions of various health and health care priorities. In 2006, for example, over 90 percent of thought leaders thought that health insurance coverage was a high or very high priority for the nation.

Finally, information from both thought leaders and grantees helps the Foundation understand whether these constituents feel that it is:

- Working on issues important to the United States;
- Making long-term commitments to important issues;
- Supporting and building leadership in health and health care.

Customer Service

The survey of grantees, conducted by the Center for Effective Philanthropy, asks how they feel the Foundation is treating them (see box). Indicators from this survey form the basis of the customer service section of the Scorecard and include questions such as these:

- Is the Foundation clear about the types of proposals it will fund?
- Is it clear in communicating its goals and objectives?
- Is it fair throughout the application process and responsive throughout the lifetime of a project?
- Are the Foundation's program officers approachable, courteous, and helpful?
- Is the technical assistance provided by or through the Foundation adequate?

The seriousness with which the Foundation takes this information is illustrated by its response to the grantee survey in 2004 that indicated its customer service was not up to that of other foundations. This led to an internal “quality improvement initiative” that resulted in an almost complete overhaul of the Foundation’s grant review and approval procedures and the way in which staff members communicated with grantees and potential grantees.

The Grantee, Thought Leader, and Public Opinion Surveys

The *grantee survey* asks approximately 300 grantees about key aspects of the Foundation’s service, programs, and impact. It contains questions on:

- Grantees’ perceptions of the Foundation’s impact on their organizations and their field;
- Whether the Foundation is addressing the most important health and health care issues and is willing to commit the time and the resources needed to achieve its goals;
- How fair the staff is in its interactions with grantees;
- Whether the grant application and reporting requirements are burdensome;
- How clearly the Foundation communicates its goals and strategies.

The survey is confidential and allows for comparisons over time. Recently, it has allowed for comparisons with other funders because of the Foundation’s participation in a survey fielded by the Center for Effective Philanthropy that elicits similar information from a number of foundations. Finally, the Foundation periodically surveys applicants it has turned down to look for warning signals coming from this group.

The *thought leader survey* interviews decision-makers from universities, health plans, hospitals, health associations, government, the media, and public health organizations. Those in this group are questioned about their knowledge of the Robert Wood Johnson Foundation as an organization and their views on its priorities, its reputation, and the quality of the information it produces. Those who are familiar with the Foundation are asked to rate its impact in addressing problems related to health and health care as well as the impact of its specific strategic areas (for example, quality of care or health insurance coverage). This survey is also confidential, and it, too, allows for comparisons over time.

The *public opinion survey* queries the public about their views on the top issues facing the nation and the Foundation’s priorities. Respondents are asked to list the issues they think the government should be addressing and to rate the American health care and public health systems. They are also asked to name the top medical care and public health concerns facing the nation. The survey then asks respondents to rank the Foundation’s areas of interest and a few other select health care issues.

Staff Satisfaction and Financial Performance

This section of the Scorecard looks first within the Foundation to determine what the staff thinks of it as a place to work (see box). This includes information about the staff’s overall satisfaction, its ratings of management, and opinion of its ability to communicate concerns up through the ranks of the Foundation.

The Staff Survey

The *staff survey*, which surveys employees anonymously, asks staff members how well they think the Foundation is doing in meeting its guiding principles—a core set of values that promote good stewardship of Foundation resources, fairness in the treatment of grantees and the field, and professional, ethical staff conduct. The staff survey asks a number of questions to determine whether staff members feel that the guiding principles are honored by the leadership of the Foundation and are useful in guiding everyday transactions. It also asks questions related to the working environment at the Foundation and the staff’s judgment of the Foundation’s program development and impact. For example, staff members are asked whether sufficient effort is made to get their opinions, whether there is an environment of teamwork at the Foundation, and whether the Foundation has clear goals and objectives.

This section of the Scorecard also reports on how the investment portfolio is doing. The indicators include the return rate and the volatility of the Foundation’s endowment portfolio.

Grants Management Performance

The Scorecard also contains, as an appendix, a review of data from the Foundation’s grants management system to see how many applications are being submitted and, ultimately, being funded; and how the work of the Foundation and its grantees is being disseminated in print and on the Web. It also contains a section that outlines changes that were and will be made.

Tier 4: Grant Results Reports and *The Robert Wood Johnson Foundation Anthology Series*

The fourth tier of the Robert Wood Johnson Foundation’s evaluative activities, which identifies and shares lessons from the Foundation’s grantmaking, takes the form of two publication vehicles: the grant results reports and the annual *Robert Wood Johnson Foundation Anthology*. Both take a broader view of evaluation and attempt to share the understanding gained from the Foundation’s investments with as wide an audience as possible.

The grant results reports now total more than 2,000 distinct reports that are posted on the Foundation’s Web site. The reports are prepared by a team of consulting writers who are asked to interview key players involved in the grant, read the written record, and prepare a report that summarizes what was actually done with the grant funds and what findings, results, and lessons learned emerged. The reports attempt to tell the stories in a manner that lets the reader come to conclusions about what the facts and experiences suggest about success and failure. The grant results reporting unit has recently started to prepare topic summaries, also posted on the Foundation’s Web site, that synthesize the key outcomes and lessons from reports on a particular topic, such as positive youth development or consumer choice in long-term care.

The grant results reports have emerged as the second most visited area of the Foundation’s Web site (just behind the section that describes how to apply for funding). The reports are read by people doing research or planning an initiative that replicates one supported by the Foundation in the past; people interested in knowing what the Foundation funds; and the Foundation’s staff members attempting to learn from past lessons to guide current grantmaking.

The Robert Wood Johnson Foundation Anthology, published annually by Jossey-Bass, examines approximately 10 topics each year. These topics may be an area of grantmaking, such as health insurance or tobacco policy, or a specific program. Or, in an effort to demystify philanthropy, they may provide an insiders' view of how the Foundation reached decisions or chose one path over another. The writers include award-winning journalists, Foundation staff members, and outside evaluators.

The authors sift through the written record, interview key players, and make site visits. They are asked to write an interesting, jargon-free chapter that lets readers know why the Foundation decided to fund the activity, what the program or programs actually did (or is doing, in the case of existing programs), what has been accomplished, and what lessons can be drawn. The book is distributed to more than 10,000 health care experts, foundation staff members and trustees, and government officials, and is available on the Foundation's Web site.

Taken as a whole, the *Anthology* series, the grant results reports, and publications that emerge from Foundation-funded evaluators (their reports are also available on the Foundation's Web site) offer an extensive record of the Foundation's successes and failures. Serving as a guide to policy-makers, health care leaders, researchers, Foundation staff members, and the general public, they represent one way in which the Foundation tries to be accountable to the public and transparent in its grantmaking.

An Assessment

The Foundation prides itself on learning from its programs and its evaluations, constantly attempting to improve how it learns. The evolution of the Foundation's evaluation—from evaluation of individual programs to assessments of the impact of portfolios of grants and the production of an internal Scorecard—demonstrates the importance of an institutional culture that promotes continual questioning and desire to learn from past experience.

Even so, the practice of evaluation, assessment, and learning still faces challenges. For example, in its traditional program evaluation, the Foundation needs to be sure that the programs we fund to test new ideas are actually designed so that a convincing test can be conducted. A culture such as that of the Robert Wood Johnson Foundation that wants to choose the best grantees and the best programs—rather than comparing a demonstration site with a control site—can actually work against learning. From a strict learning perspective, sites should be chosen so that a legitimate comparison can be made between places that try an idea and places that do “normal practice.”

Another challenge is rationing evaluation dollars so they are spent on cases for which testing and learning are possible. Evaluations are often expensive, and confidentiality requirements and the difficulty of collecting survey data have increased their cost in recent years. When a program is not designed to test a new idea, no evaluation may be needed. The decision *not* to evaluate needs to be made more frequently to have money available for more comprehensive evaluations of programs where testing is possible.

Determining whether a set of Foundation investments have had a causal effect remains a challenge. Currently, performance assessment measures correlation more than causation. Correlation can imply causality when the Foundation's investments are significant enough to be the only likely cause of

some event or change. But when the Foundation attempts to affect a complex social situation—such as reducing smoking or improving the quality of health care—it is difficult to know the extent to which the Foundation is responsible for the improvements or whether other factors are the cause.

It is insufficient simply to say that when something positive happens related to a grant program that it is because of the Foundation or its grantees. This expanded sense of impact can be spurious at best. To assess success or failure, the program staff members must be able to articulate the link between the strategy and the expected short-, medium-, and long-term outcomes. Without this type of roadmap, clear assessments will never emerge.

Similarly, the durability of strategy must be considered. On the one hand, it is important for strategy to be flexible enough to react to changes in the environment. On the other hand, if strategies and indicators of success or failure change frequently, then performance assessment is not possible. Constant changes in direction—as opposed to the fine-tuning of a set strategy—generally indicate that a strategy is not working, is not being executed effectively, or was misguided from the start.

Finally, efforts to judge whether the organization itself is working efficiently and effectively require a blend of qualitative assessment and quantitative assessment. The current Robert Wood Johnson Foundation approach perhaps relies too much on quantitative indicators. If the Foundation really wants to know how it is doing, it may need to take the approach of former New York City Mayor Ed Koch and ask in plain language, “How’re we doing?” There are many Robert Wood Johnson Foundation-watchers across the country; frank conversations with them could round out the information that the thought leaders’ survey provides. Such qualitative information could add an important dimension for the Foundation’s Board of Trustees and senior staff to consider when assessing organizational effectiveness.

The path to developing more effective ways to track and assess performance is the same one the Foundation advises for improving health care quality: continuous quality improvement. Constructing a viable evaluation strategy for a philanthropy is not a one-time building project. It takes constant attention, tinkering, and questioning. While the Robert Wood Johnson Foundation can take pride in its leadership in the field of philanthropic evaluation, it needs to learn from its peers, understand emerging trends in performance assessment, and remain open to evolution in its approaches.

Notes

1. Lavizzo-Mourey R. "Foreword." In *To Improve Health and Health Care, Vol. VII: The Robert Wood Johnson Foundation Anthology*. San Francisco: Jossey-Bass, 2004.
2. Wielawski I. "Fighting Back." In *To Improve Health and Health Care, Vol. VII: The Robert Wood Johnson Foundation Anthology*. San Francisco: Jossey-Bass, 2004.
3. Colby DC. "Building Research Capacity in the Sciences." In *To Improve Health and Health Care, Vol. VI: The Robert Wood Johnson Foundation Anthology*. San Francisco: Jossey-Bass, 2003.
4. Brown LD and McLaughlin C. "Constraining Costs at the Community Level: A Critique." *Health Affairs*, 9(4): 5–28, 1990.
5. Kaplan RS and Norton DP. *The Balanced Scorecard: Translating Strategy in Action*. Cambridge, Mass.: Harvard Business School Press, 1996.