Measuring the Impact of Social Programs: A Review of Best Practices
Lesley Sept, Sandra Naylor, Randy Weston
February 9, 2011

Introduction

The question of how to measure the impact of a global corporation’s social efforts can seem almost unmanageable. The facets of influence seem to multiply the harder you look at the question. Who is being impacted? (Individuals, communities, an industry sector, a nation, the world?) What’s the nature of the impact? (Something we can easily measure, like greater income, or something more nebulous like empowerment?) Finally, how efficiently is that impact delivered? (In terms of management, operations, return on investment? How do our efforts leverage the firm’s core competencies or unique resources? How did we do against some measure of what impact we might have achieved had we invested our resources differently?)

Evaluation can range from collecting a few key data points every month to undertaking a decade-long academic analysis. Which is better? It all depends on what your goals are.

For example, take a PepsiCo water resources program in Ghana. When the program was in the pilot stage, the team developed and used a few key performance indicators to track throughout the pilot. Their goal was to assess the pilot’s success as well as get information to refine the service delivery approach in real time. Keeping an eagle eye on a handful of indicators is a great way to focus on key performance issues and adapt quickly. For instance, one key metric was: How far are people traveling to use the new water source? When program staff noticed that people in remote areas were not using the well, they dispatched a truck to deliver to outlying communities. In this case, evaluation was embedded throughout the program. Data collection was no more formal than a man standing by a spigot for 30 minutes, asking, “How far did you travel to get here today?” The results could be communicated by text message.

On the opposite extreme, successful evaluation can take the form of rigorous academic studies conducted over the course of months or years. For instance, consider a formal impact evaluation such as the one Stanford University is conducting for a healthcare logistics improvement project in Africa called Riders for Health. A quasi-experimental time-series design is being utilized in order to isolate programmatic impacts. Over five years, researchers will compare districts receiving program services with ones that do not to gather proof of program impact. The goal is to contribute to the body of knowledge on what works in improving health care delivery in rural parts of developing countries, to support future funding decisions.
In this paper we will review best practices for assessing social programs. First we will provide a high-level overview of common methods used to evaluate social impact. We will offer a comprehensive taxonomy of KPIs used to track social impact and offer guidance on selecting appropriate indicators. A summary of the key methodologies for integrating cost into estimations of social value creation in order to assess social return on investment will be provided. Throughout, we will provide case studies illustrating these concepts in use.

1. Overview of Social Impact Evaluation Methods

The first step in choosing an evaluation method is to clarify what questions need to be answered. The goals of the evaluation, as well as the target audience for the findings and the resources available for the evaluation, are all important factors in selecting an assessment method. At a high level, there are four primary ways to evaluate a program. These include impact evaluation, performance measurement, process evaluation, and social return on investment evaluation. Here we will briefly introduce each method before diving deeper into some of the most relevant evaluation designs:

- **Impact Evaluation.** Focuses on questions of causality. Did the program have its intended effects? To answer this question, impact evaluations compare program outcomes with some measure of what would have happened without the program. Generally speaking, this method offers the most rigorous assessment of program effectiveness, but requires more time, resources and expert assistance than other methods.

- **Performance Monitoring.** Provides information about how key aspects of a program are operating and the extent to which specified program objectives are being attained. Used by service providers, funders, and policymakers to assess the program’s performance and accomplishments. Typically this method offers a cheaper, faster way of gathering data on a program’s on-going performance.

- **Process Evaluation.** Answers questions about how a program operates and documents the procedures and activities undertaken in service delivery. Useful to practitioners in replicating program strategies, as well as in identifying problems or adapting programs.

- **Social Return on Investment Evaluation.** Incorporates information on the cost of a program, often in relation to alternative uses of the same resources and to the benefits produced by the program. This method makes significant, often prohibitive, demands on an organization’s time and resources.

---

1 In this section, we draw heavily from the Urban Institute’s “Evaluation Strategies for Human Services Programs: A Guide for Policymakers and Providers,” 1996, by A. Harrell et al. This article can be found online at: http://www.urban.org/publications/306619.html
1.1) Impact evaluation

Impact evaluation methods compare program outcomes with some measure of what would have happened without the program, as noted above. This approach is especially important in the early stages of a program’s life, when it is not yet clear how effective the program’s strategy is at creating the desired social impact.

There are three common impact evaluation designs: experimental, quasi-experimental and non-experimental. Experimental designs are the most powerful and produce the strongest evidence. It is not always possible to use an experimental method, however. In those cases, one of the other two designs must be chosen.

While impact evaluation approaches tend to be carefully designed for each program, it is possible to give an overview of the structure, advantages and disadvantages of each common design:

- **Experimental designs.** This method is considered the gold standard of impact evaluation. Experiments require that individuals or groups, such as farmers or women’s associations, be assigned at random (by the flip of a coin or equivalent randomizing procedure) to two groups prior to the start of services. The ‘treatment’ group will be designated to receive particular services designed to achieve clearly specified outcomes. A ‘control’ group receives no services. The treatment group outcomes are compared to control group outcomes to estimate impact. Because chance alone determines who receives the program services, the groups can be assumed to be similar on all characteristics that might affect the outcome measures except the program. Any differences between treatment and control groups, therefore, can be attributed with confidence to the impacts of the program.

  Design variations include: the creation of more than one treatment group receiving different interventions, in order to estimate the relative impact of different program strategies; random selection of periods of time (e.g. offering new services on randomly chosen weeks or days) rather than individuals or groups; or a ‘staggered start’ approach where a treatment group receives services earlier than the control group.

A key advantage of an experimental design is the quality of evidence produced.

Disadvantages include the difficulty or impossibility of implementing random assignment in some cases, as well as practical concerns surrounding the cost and rigorous requirements of implementation. Some programs are inherently impossible to study through randomized experiments – for example, enforcing a law only against certain members of the population. Random assignment may be considered unethical if a program believed to be highly beneficial is denied to certain people. On a practical level, experimental designs run the most risk of being
contaminated because of deliberate or accidental mistakes made in the field. Noncomparable data may be available for treatment and control groups – often, it is more difficult to collect data from the control group than from program participants. Finally, experimental evaluations are costly, requiring expert assistance with evaluation design and monitoring as well as significant staff resources to collect data.

- **Quasi-experimental designs.** Like experiments, quasi-experimental evaluations compare outcomes from program participants to outcomes for comparison groups that do not receive program services. The critical difference is that the decisions who receives the program is not random. Comparison groups are made up of members of the target population as similar as possible to program participants on factors that could affect the selected outcomes to be observed. Multivariate statistical techniques are then used to control for remaining differences between the groups.

The advantage of quasi-experimental designs is that they can be used when random assignment is too costly or impossible to implement. Often, evaluators use existing population groups for comparison – for example, those who are enrolled in the same school in a different classroom, or attended the same school with the same teacher in the previous year.

Key disadvantages of quasi-experimental designs include their less defensible methodology and the corresponding need to be vigilant against threats to the evaluation’s validity. There are two key threats to validity. First, matches based on similarities at a single point in time do not always result in groups who are comparable over time. Groups may become increasingly different over time independent of the program. Second, differences in variables not used in matching may have a substantial effect independently of the program being evaluated.

- **Non-experimental designs.** These methods track changes in outcomes among program participants, but do not include comparison groups of other individuals or groups not exposed to the program.

There are four main types of non-experimental designs: before and after comparisons, time series designs, cross-sectional comparisons, and panel designs. Using aggregate data, before and after comparisons measure outcomes prior to and upon completion of an intervention, inferring impact from these differences. Time series designs are an extension of before and after comparisons, with added data collected while a program is in progress. Cross-sectional comparisons use individual-level data to survey participants after a program’s completion to get a sense of correlations between outcomes and the duration, type and intensity of services received. Panel designs track individual outcomes before, during and after a program, to identify different patterns of change associated with individual characteristics of participants and control for other events to which they were exposed.
Non-experimental designs have several practical advantages. They are relatively easy and inexpensive to conduct. Individual data for cross-sectional or panel analysis is often collected routinely by program staff. Aggregate statistics may be obtained from the program or other community or public agencies. For example, relevant economic or social indicators may be collected at the local, state or national level.

The primary disadvantage of non-experimental designs is, of course, the inability to demonstrate conclusively that any changes in participant outcomes were due to the program.

For more information on how to conduct an impact evaluation, refer to:


Summary:

- When seeking proof of a program’s impact, consider impact evaluation methods. Results of these rigorous analyses have high credibility and so create higher visibility when seeking to publicize the results of an organization’s efforts.
- Keep an eye out for opportunities to randomize: to compare a program’s results against a control group. For example, with minimal effort and resources it would be possible to randomly assign women’s groups to receive or not receive the Shea Nut program (all would eventually receive it).
- The downside of impact evaluation is that it takes a long time and can be expensive. Also, impact evaluation cannot be used to directly compare programs. When more timely information is needed or when program results need to be compared or aggregated, consider performance monitoring or social return on investment methods. These two approaches are discussed in the sections that follow.

2. Performance Monitoring and Benchmark Creation

While in-depth impact evaluation provides the most conclusive demonstration of a program’s social impact, alternative methods exist when it is more important to get a timely, low-cost sense of how a program is performing. Key performance indicators (KPIs) can be defined, collected and assessed on an ongoing basis to evaluate a program’s operations and to estimate its social impact. A set of KPIs may be
developed for an individual program, or for a collection of programs through the use of standardized benchmarks.

### 2.1) Questions performance monitoring can answer

Performance monitoring involves defining a few important indicators of success and tracking how these metrics change over time. As in a for-profit setting, these metrics can be financial or operational, in addition to more traditional social sector metrics following outcomes of program participants. Performance monitoring is well-suited for answering questions like:

- How are key aspects of a system or program operating? For example: What is the on-time percentage of medical supply deliveries to a health clinic this quarter?
- Are pre-specified program objectives being attained, and to what extent? For example: How have the incomes of Ghana shea farmers in the treatment group changed this quarter?
- How satisfied are those being served by the program? Such information is typically gathered through customer surveys.
- How efficient, effective and productive is the program, when we compare inputs with output and outcome indicators? For example: What is the cost per child of an after-school tutoring program? Of a 1-point increase in grade point average (GPA)?
- Are there any failures to produce expected program outputs? This information can be inferred from the types of questions listed above, and can be used to manage or redesign program operations.

Unlike impact evaluation, performance monitoring does not involve a rigorous effort to determine if observed changes were due to the program or other, external factors. Any changes in participant outcomes can only be assumed to have occurred due to the program’s interventions.

For an example of performance monitoring in use, see the PepsiCo case study on the following page.
Case study: PepsiCo and Safe Water Network

Water plays a major role in PepsiCo’s business and, for the past few years, it has played a central role in PepsiCo’s social and environmental sustainability strategy. The PepsiCo Foundation partnered with the Safe Water Network (SWN) to implement safe water programs in communities in Ghana, India and Bangladesh, committing $3.5 million over three years.

To develop successful pilot programs and scale them up, the project uses metrics at each step along the way. SWN begins with a thorough market assessment of the needs of the community. Working with Johns Hopkins University, as well as government organizations and other NGOs, the organization quantifies the community’s current conditions. With this baseline, SWN can not only plan a pilot that fits the community’s needs, but also set targets for improvement and measure success.

Armed with this information, the organizations plans a pilot project. Throughout the pilot planning and implementation phases, everyone is looking for ways to continuously improve efforts. Rigorous tracking of performance metrics facilitates learning.

For example, in Ghana people living within 500 meters of a water facility are the largest adopters, but people one to two kilometers away from these community centers are also a target community. The initial assessment suggested that locating plants in commercial centers serving surrounding villages and towns would be optimal. But SWN vice president Amanda Gimble said it was clear this wasn’t enough, with many people living too far from community centers to make the trip to get the clean water. How to serve them? “We put into place a delivery system in which we delivered water by truck to these areas and overnight the adoption levels doubled,” she said. “We demonstrated very quickly that people are willing to pay for water being transported to their home. Those are the kinds of lessons we are using to shape future programs.”

These sorts of lessons come from the organization’s rigorous assessment of its impact. Water use is monitored from the time it is dispensed until it is used, monitoring quality all along. The lessons learned are then shared not only with partners but with other organizations working on water quality issues, so everyone can benefit from the lessons.

2.2) Common KPIs

While social programs vary widely in terms of their goals and activities, it is possible to think about their results in four broad areas: program outcomes, participant outcomes, community outcomes, and organizational outcomes.

Program indicators track the activities, resources and results of service delivery, including number of people served and program quality measures. Examples of participant outcomes include measures of participant health or income. Community outcomes include the program’s broader impact on society at large, e.g. environmental impact. Finally, organizational outcomes encompass financial, management and governance indicators.

The Appendix provides an overview of common indicators in each of these four areas. While it would be impossible to list all potential indicators, the Appendix shows a comprehensive list of outcome areas worth considering in drawing up a set of KPIs.

2.3) Selecting key performance indicators (KPIs)

The most common way to select appropriate KPIs for a program is to use a logic model. A logic model is essentially a picture of how an organization does its work. Figure 1 shows a basic model. Logic models link results (short-term and long-term) with
program activities and processes, and illustrate the theoretical assumptions of the program — its so-called “theory of change.” In short, a logic model is an illustration of how a program’s day-to-day actions and resources are believed to create social impact.²

![Logic Model Diagram]

Figure 1. The Basic Logic Model.

There are three steps to follow in developing KPIs for a program. First, create a logic model. Then identify relevant indicators at each stage of the model. Finally, carefully select from the many potential indicators to develop a short list of key performance indicators.

**Step One: Create a Logic Model**

The first step is to understand how a program operates and its expected social impact by building a logic model. Stakeholders, including program leadership and funders, often collaborate in this process to create a shared understanding of program activities and expectations.

Let’s begin by defining the terms used in a logic model:

- **Resources, or inputs,** include the human, financial, organizational and community resources that a program draws upon to do its work.
- **Activities** include the processes, tools, events, technology and actions that an organization implements using its resources.

Resources and activities combine to create desired results. Desired results can be measured across three time frames:

- First, in the short term, a program’s direct effects are called **outputs**: the types, levels and targets of services delivered.

• Over the next several years, a successful program will see outcomes: specific changes in participants’ behavior, knowledge, skills, status or level of functioning.
• Finally, impact describes any fundamental intended or unintended changes in organizations, communities or systems. Long-term impact may take up to a decade to assess.

Note that our discussion here is simplified, in order to offer a high-level overview of logic modeling. Organizations often create sophisticated models to better understand and evaluate their theory of change. Additional resources for developing a logic model are provided at the end of this section.

Once a model exists that does a reasonable job of illustrating how a program intends to create desired outcomes, it is time to move on to the next step.

**Step Two: Identify Relevant Indicators at Each Stage of the Program**

By mapping out how a program works in a logic model, it is possible to identify relevant performance indicators at each stage of the model. As an example, see on the following page some candidate indicators for a Shea Nut Reinforcement project in Ghana that utilizes microfinance, education, and technology to improve the incomes and living conditions of women who pick sheat nuts and women who process nuts into shea butter.
Case Study: SAP and PlaNet Finance’s shea project in Ghana

For this project with female shea nut farmers in Ghana, here are some potential indicators at each project stage.

<table>
<thead>
<tr>
<th>Resources/Inputs</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>• SAP’s funding</td>
<td>• Number of women eligible for the program</td>
<td>• Increase in program participants’ shea product income by end of program</td>
<td>• Percent of participants using trees as firewood</td>
<td>• Women’s empowerment metric, e.g. average income, health, education, in program’s regions or throughout Ghana</td>
</tr>
<tr>
<td>• Number of women groups or clusters formed</td>
<td>• Number of training sessions offered</td>
<td>• Number of women who complete business training</td>
<td>• Percent of participants selling key household assets during lean season</td>
<td>• Price, quality and/or total shea production in program’s regions, Ghana or globally</td>
</tr>
<tr>
<td>• Capital available for loans for shea nuts and butter activities</td>
<td>• Number of communities with services available</td>
<td>• Percent of participants who opt to take a loan</td>
<td>• Percent of participants using trees as firewood</td>
<td>• Metric tracking supply chain consistency/reliability in program’s regions, Ghana or globally</td>
</tr>
<tr>
<td>• Full-time-equivalent staff available</td>
<td>• Quantities of equipment delivered to enhance business activities, e.g., Tarps to dry nuts, protective gear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Number and average dollar value of loans made</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lending metrics: repayment rate, interest rate, duration of payback period</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In creating a list of metrics, keep in mind that there are certain basic criteria for high-quality indicators. The Appendix of this document offers a list of common performance indicators. Whether selecting off this list or creating your own, indicators should be:

- Relevant: Measuring important dimensions of program performance, appropriate, predictive, timely. Metrics should be clearly attributable to program activities.
- Observable: Practical, cost effective to collect, measurable.
- Understandable: Make sense to evaluators and stakeholders.
- Specific: Unique, unambiguous.
- Time-bound: Covering a specified period of time, e.g. a quarter or year.
• Valid: Providing reliable, accurate, unbiased, consistent and verifiable data. Can be operationally defined in a straightforward manner so data can be collected consistently over time, across data gatherers, and across geographies.

**Step Three: Develop a Short List of Key Performance Indicators**

Once a logic model exists, many potential indicators will have been identified at each step of the logic model. How do you distill this list into a set of key performance indicators? A set of KPIs for a program should cover all important steps in the model, while being realistic about the burden of data collection and boiling down the KPI list to as small a number of metrics as possible. The Urban Institute suggests that three to seven KPIs are generally sufficient for performance monitoring.3

In selecting a short list of KPIs, consider the following best practices:

• Comprehensiveness: Covering inputs, outputs, and service quality measures, as well as longer-term outcomes and relevant items of customer feedback.
• Mutual exclusivity: Uniquely relevant indicators. Minimize redundancy.
• Logical consistency: Align indicators with objectives and program activities in organization’s mission statement. Consider the extent to which the program has influence/control over outputs or outcomes.
• Keep it simple: Limit data collection to a few key indicators, to facilitate management use of data and minimize collection costs.
• Measure regularly: Frequent and timely measurement, ideally on a quarterly basis, will allow shifts in program operations to be tracked. Of course, data quality should not be sacrificed for the sake of timeliness. Evaluation plans should include adequate support and plans for training and technical assistance, as well as quality control.

For more resources on how to develop a logic model, see:


**2.4) Benchmarking**

Having identified KPIs to track at the program level, a natural next step is to think about developing benchmarks to compare similar programs. Common benchmarks would allow funders to aggregate their social impact across programs. Shared standards could also funders to compare the relative effectiveness of programs.

3 Urban Institute and the Center for What Works, Outcome Indicators Project: [http://www.urban.org/center/cnp/projects/outcomeindicators.cfm](http://www.urban.org/center/cnp/projects/outcomeindicators.cfm)
Case study: Common Reporting Standards - Pulse and IRIS

Two innovative efforts to assess organizational performance in the social sector are being led by the Acumen Fund, the Rockefeller Foundation, and others. These funders seek to develop comparative performance systems that promote learning and help increase knowledge at the sector level by allowing users to benchmark and compare organizational outcomes using identical indicators and data collection methods.

Two interrelated initiatives are being pursued by these groups. IRIS is a set of data collection standards to facilitate collection of comparable information, while Pulse is an online data management system that supports the gathering of data – especially data that conform to IRIS standards.

More specifically, Pulse is a web-based data management system originally developed by the Acumen Fund for use by portfolio managers. It allows managers to choose from a pool of general and sector-specific indicators (or create their own); to track financial, social, operational and environmental metrics; and run longitudinal and comparative reports. Pulse recently became available to the public. Licenses are available free of charge for nonprofits. For more information see Pulse’s website:
http://www1.app-x.com/pulse

Impact Reporting and Investment Standards (IRIS) is an emerging open-source reporting framework that will allow users to define, track and report the performance of impact investing capital, and to benchmark performance metrics at the portfolio and sector levels. The Rockefeller Foundation, B Lab, and the Acumen Fund are leading the effort to develop IRIS. Pulse uses the IRIS taxonomy of indicators. Because of the focus on investment performance, the IRIS taxonomy primarily includes financial indicators, not more comprehensive social impact assessment metrics. More information on IRIS is available here:
http://iris.thegiin.org/

To some extent, KPIs will inevitably vary across programs, even those in the same sector. Differences in program activities, client populations, and so on will lead to diversity in appropriate KPIs. That said, funders are increasingly interested in creating benchmarks to assess performance – seeking to understand their own performance as an organization as well as that of their grantees.

One way for a funder to develop a set of common benchmarks would be to consider its own mission and determine if certain universal outcome measures are appropriate to evaluate the funder’s success in achieving its vision. For example, IT capacity building programs may operate in various geographies, targeting various populations using a variety of strategies. If a funder is thoughtful about the organization’s goals, it could be that one or two metrics – e.g. changes in employment or business ownership among program participants – can be used as a bottom-line KPI across many programs.

Of the four outcome areas discussed above – program, participant, community and organizational – the area that best lends itself to benchmarking is organizational performance. Financial, operational and governance metrics can be tracked across organizations in virtually any sector. Such metrics are also more closely related to those used in the for-profit arena. One example of efforts to create a benchmarking system using organizational indicators, for the most part, is described in the Pulse and IRIS case study (see sidebar at left).

While common KPI benchmarks may make sense at the sector level, another way to consider measuring diverse programs on a common yardstick is to assess their social return on dollars invested. Incorporating measures of cost could allow funders to compare
apples and oranges, the way investors in for-profit businesses can compare a manufacturing firm with a retailer by considering bottom-line profitability metrics. We turn next to a discussion of how to measure social return on investment.

Summary:

- Tracking key performance indicators is an effective way to get timely, useful data on how a program – or several programs with a similar big-picture goal – are doing.
- KPIs should be thoughtfully chosen: keep the list small (approximately 5 indicators) and closely aligned with the program or funding organization’s mission.
- Despite the wide variety of social programs in existence, it’s possible to track them using common measures of program, participant, community and organizational outcomes.

3. Review of Social Return on Investment Frameworks

Measures of social return on investment, or social ROI, offer the promise of allowing funders to compare social programs using a common yardstick. Efforts to incorporate cost into measures of social impact in order to estimate social value creation are still nascent. Despite significant work to build robust methodologies, problems remain with the high cost of social ROI estimation efforts and the inherently unquantifiable nature of some types of social impact.

In this section, we will provide an overview of two ‘classical’ methods found in textbooks, cost-effectiveness analysis and cost-benefit analysis, as well as two emerging frameworks developed in the past 20 years by practitioners – social return on investment (SROI) and expected return. We will offer a short description of each approach, discussing the contexts in which it is often used and highlighting the method’s relative advantages and disadvantages. References will be provided for more technical information on how to calculate each measure. Finally, we will offer a list of best practices in estimating social ROI.

3.1) Cost-Effectiveness Analysis

Cost-effectiveness analysis (CEA) involves the calculation of a ratio of cost to a non-monetary benefit or outcome (e.g., cost per high school graduate, cost per person treated for malaria). The purpose of CEA is to combine appropriate measures of program outcomes with cost, in order to rank and select among program alternatives. CEA allows you to sidestep the uncertainties about how to value program benefits by looking at the ratio of benefits to costs without reducing them to common units (i.e., dollars). The approach is used in the private, public and nonprofit sectors, particularly in health care, where the costs

---

Case Study: Benefit-Cost Ratios at the Robin Hood Foundation

The Robin Hood Foundation is an organization whose sole aim is to reduce poverty in New York City. The organization provides ongoing grants to more than 200 NYC nonprofits that operate in four areas: Jobs & Economic Security, Education, Early Childhood & Youth, and Survival.

To evaluate these diverse programs, Robin Hood developed a Benefit-Cost Ratio method. The purpose of the method is to translate the performance of each grant into a single, monetized value that estimates the collective benefit to poor individuals generated by Robin Hood’s investment. Whether evaluating health clinics, microlending projects or after-school programs, staff make their best effort to quantify the benefits created and compare this with program cost. The goal is to answer the question, “Which programs should we fund and how much should we spend on each?”

In practice, Robin Hood applies this method when considering funding a new program and when evaluating programs on an on-going basis. Program officers are required to calculate Benefit-Cost Ratios for all of their new and renewal program funding proposals on an annual basis. While the foundation operates in four areas, there are no fixed allocations to each area (e.g., $1 million for Education and $2 million for Survival). Rather, the foundation seeks the highest-impact projects from any sector, and lets the allocations develop passively as a result of which individual grants are selected.

For more information on how Robin Hood calculates Benefit-Cost Ratios, see:


of an intervention are compared to its impact on an individual’s quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs).

One advantage of CEA is that the method allows the comparison of program alternatives within the same sector, so long as they are seeking the same outcome. Another key benefit is that it can be used in situations when monetizing the benefits of a program is not possible or appropriate.

The primary disadvantage is that CEA does not allow for meaningful comparison of programs across sectors, or of programs in the same sector that seek significantly different outcomes.

3.2) Cost-Benefit Analysis

Cost-benefit analysis (CBA) monetizes the benefits and costs associated with an intervention and then compares them to see which one is greater. This is the most demanding approach to analyzing costs and outcomes, because it requires a comprehensive measure of program impacts (e.g., direct and indirect, primary and secondary, tangible and intangible impacts) across stakeholders and the ability to put a dollar value on those impacts. This comprehensiveness allows CBA to offer a full accounting of a program’s net benefits to society as a whole, as well as to each of its stakeholders.

As with CEA, CBA allows for the comparison of program alternatives to determine which is most effective. However, CBA also provides a bottom-line assessment of whether a program is ‘worth it’ – whether the program generates more benefits to society than the resources expended to implement it.

An example of CBA used in practice is at the Robin Hood Foundation, an organization focused on poverty in New York City (see case study at left).
Besides benefit-cost ratios, the output from CBA analysis can include: measures of net benefits (benefits minus costs) also known as Net Present Value (NPV), or the Internal Rate of Return (IRR) - an estimate of the annual social benefit return of a project. CBA is widely used in the private, public, and increasingly the nonprofit sectors to help decision-makers prioritize or decide among various uses of funds for programs and projects.

CBA’s key advantage is that it allows for comparisons of programs with varying strategies in a wide variety of sectors.

The disadvantage of the approach is that it is highly demanding to gather data across various dimensions of social benefit accruing to various stakeholders. Moreover, once gathered, the data on social outcomes needs to be monetized in some fashion.

3.3) Social Return on Investment

Social Return on Investment (SROI) analysis involves measuring the non-financial impact of an investment. The method offers guidelines to monetize some aspects of social value creation, especially economic metrics like increases in participant income due to education or job placement. Best practices in SROI analysis also include gathering information on benefits that cannot be quantified, to permit a sophisticated understanding of impact.

The SROI methodology was originally developed by the venture philanthropy nonprofit REDF. While that organization ultimately ceased to use SROI as a basis for funding decisions or program evaluation, REDF and others have continued to develop SROI approaches.

At heart, SROI analysis involves monetizing the social and environmental benefits of a social program – or the operations of a business – and calculating the return on investment as an investor would. For a discussion of the technical details and difficulties of this method (for example, what is the appropriate discount rate to use in such a calculation?), see the resources listed at the end of this section.

Advantages of SROI include the ability to quantify the more tangible aspects of social value creation. Also, the method’s similarity to traditional for-profit calculations of return on investment may make it a good fit for companies looking to augment their financial performance metrics with information on social impact.

Disadvantages include the long time horizon required to truly estimate social impact – often, five years or more. There is also a risk that too much attention will be placed on the bottom-line SROI number, when advocates emphasize that this is an evaluation process that requires an understanding of the ‘unmeasurables’ that influence interpretation of the number. It is worth noting that REDF, the organization the originated the SROI methodology, ultimately opted not to use it for funding decisions.
or evaluation purposes. In their experience, implementing SROI required extensive time and resources and involved rather subjective quantifications of impact. Moreover, the unquantifiable impacts were often much more significant in assessing program success than that which could be measured.

3.4) Expected Return
Unlike the other measures discussed so far, the William and Flora Hewlett Foundation’s expected return methodology involves estimating the potential impact of a grant, rather than evaluating after the fact. Developed in 2007, the method involves considering the best-case scenario of program impact, then adjusting this ideal performance estimate by the odds of success. See a summary of the ER calculation in the figure below. To determine the value of each variable, funders engage in a process of quantifying the organization’s goals and using existing research and information to estimate a project’s contribution toward those goals.

![Figure 1](image1.png)

The purpose of calculating expected return is two-fold. Before funding, the ER estimation process requires funders to think critically about a program’s theory of change and do their homework on how similar programs have fared – that is, to research best practices. After grants have been made, initial assumptions can be compared to the data coming in on program performance.

The key advantage of ER is that the method offers guidance at perhaps the most important moment – the time of investment. ER ensures that program investments are consistent with the funder’s mission and the latest research on program effectiveness.

Disadvantages include the lack of rigor in quantification. Estimates of input variables like likelihood of success may amount to little more than a guess. ER also cannot be used after the fact to measure social return on investment or comprehensively evaluate program success.

3.5) Best practices in integrating cost
The process of gathering robust data on social impact can soak up a lot of time and money over the course of a decade. Among practitioners, some have abandoned attempts to rigorously estimate social ROI. Those who have met with more success using the method have followed some of the following best practices:

**Accept that there is no ‘silver bullet.’** There is no one special metric that can tell you if a project is a success or a failure. There has been a very strong desire among some in the funding community to find a bottom line number akin to profit or loss in a business, which would reduce the role of subjective judgment in grantmaking. This desire may never be realized in the social sector due to the difficulties of measuring social benefit and a multivariate, diffuse definition of what constitutes success – as opposed to a business’s ultimate aim of profit.

**Apply one method consistently throughout an organization.** The value of estimating social ROI comes not from any great precision in the results, but rather, from the ability to compare the results across programs to drive decision-making. Organizations like Robin Hood and the Hewlett Foundation have derived more value from consistent evaluation processes that allowed relative program benefits to be assessed, rather than rigorous attempts to generate absolute measures of program impact.

**Consider carefully the decision to integrate cost.** What, exactly, is the practical benefit of including cost information in evaluations of program impact? Some have found that it is more useful to focus only on social benefits when looking at programs across various sectors, strategies and program sizes, without trying to compare such a diversity of programs using social return on investment methods.

For more technical information on how to calculate these social return on investment measures, including formulas and a discussion of more technical issues in measurement, see:

- Social Return on Investment: A rich collection of resources from REDF, the originator of social ROI methodology:

**Summary:**

- Using social ROI methods to compare programs is a potent tool when programs have the same ultimate goal, e.g., getting kids to college or reducing poverty. It is also powerful when programs operate directly and independently to create desired benefits.
• Conversely, social ROI methods are less powerful when comparing very different programs, or when evaluating organizations that must collaborate with many actors or seek to achieve diffuse goals (e.g. policy change or empowerment of a group).
• Calculating social return on investment is more feasible when programs lead to greater income or provide goods and services that can be easily valued, such as job training or emergency relief.
• Practically speaking, some have become disillusioned with the time- and resource-intensive nature of conducting social ROI analysis. Incorporating cost may not always be feasible or important in making initial or follow-on funding decisions.

4. Conclusion: Toward an Integrated Approach

When a company seeks to evaluate its impact on society, there are two lenses that can be used to assess the organization’s efforts. First, the company can consider the impact of its corporate social responsibility (CSR) function, by appraising its financial and in-kind contributions in much the same way that a philanthropic foundation, government, or other donor organization would. Alternatively, a company can consider the social impact of its primary business operations as well as CSR. This ‘integrated’ approach would involve tracking the organization’s impact on customers, suppliers, employees and the community, using many of the approaches and metrics we have discussed in this paper.

Taking an integrated approach could redefine the mission and role of CSR departments within large organizations. When social impact KPIs sit alongside traditional operational considerations in business reporting and manager evaluation, social responsibility becomes a core business issue. CSR groups today typically manage financial and in-kind contributions that are often loosely related to the company’s core operations. As companies take a more integrated approach to assessing their social and environmental impact, the CSR group could evolve to develop and track the key impact metrics used by the organization, as well as driving innovation by fostering changes in processes and products that lead to greater social impact.
What might an integrated approach to assessing social impact look like in practice? In this section, we provide case studies and best practices to illustrate this vision. Among early adopters, the following approaches have led to success.

**Manage social responsibility as a core operational issue.** Core managers overseeing the supply chain or individual business units can ‘own’ social responsibility and tackle it as aggressively as they do cost, quality, speed and dependability. Operating executives can be held accountable for social impact by tracking their performance against predetermined KPIs.

For a case study describing how social responsibility can be integrated into managers’ performance mandates, see the Nike sidebar at left. Another example comes from the Brazilian health and beauty product firm Natura. This publicly-traded company tracks its contributions to the social, environmental and economic well-being of communities served on each department’s profit and loss (P&L), as well as in its corporate investor relations publications. The firm sets annual company-wide goals for supplier, customer, and environmental impact that executives are held accountable for meeting.

While improvements to internal operations are a good starting point, leading edge companies are going beyond that to identify opportunities upstream and downstream.

**Coordinate with the extended supply chain.** Often, an internal operation can achieve only limited social responsibility improvements on its own. Engaging the entire supply chain – including direct suppliers and customers, suppliers’ suppliers and customers’ customers, and competitors – can lead to breakthroughs.
and reduce risks. Industry collaboration reduces companies’ individual burdens in promoting socially and environmentally responsible practices globally. The use of key performance metrics to track performance across several organizations and stages of the supply chain is often critical. See the Starbucks case study sidebar for an example of collaboration in action.

**Develop organization- and industry-wide key performance metrics.** The tradeoff between customized metrics and universal benchmarks is as evident in evaluating companies as it is in assessing social programs. As discussed earlier, there can be a compelling rationale for shaping unique evaluation strategies for a social program or for measuring several programs against the same yardstick, depending on the goals of the evaluation.

Within a for-profit company, however, there is a strong case for centralizing evaluation efforts and using common metrics across a company’s divisions and, over time, throughout an industry. Building strategic partnerships with different partners in the value chain often requires the use of standardized metrics to ensure results are being achieved and recognized.

Building social responsibility evaluation into an organization’s core reporting and accountability systems is a complex task, but one with much greater potential impact on society than CSR departments can accomplish alone. Early adopters are finding that an integrated approach can help drive innovation, lower costs, and improve supply chain reliability. But processes, products and relationships with other organizations – the very nature of how things get done – often needs to be rethought to enable significant change.

For more discussion of these issues, see:

- Several more case studies, including Procter and Gamble, Merck, the Clinton Global Initiative and others, are also available upon request.
Appendix

As noted in section 2.2 above, there are four broad areas of potential impact: program outcomes, participant outcomes, community outcomes, and organizational outcomes. The taxonomy provided on the following pages offers a comprehensive list of a program’s potential spheres of influence, along with a common indicator in each category. For more indicators in each category, as well as sector-specific indicators, refer to the Urban Institute’s Outcome Indicators project and the Impact Reporting and Investment Standards (IRIS):

http://www.urban.org//center//cnp//projects//outcomeindicators.cfm