ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

J-PAL Cost-Effectiveness Methodology
March 30, 2010

Executive Summary

The cost-effectiveness analysis presented on J-PAL's website shows the impact that a specific policy can achieve for a given expenditure (e.g. $100). All the impact estimates are based on evidence from rigorous randomized evaluations. A number of assumptions were made in the analysis about how to measure costs and benefits. In most cases there is no single "right" answer—different approaches are appropriate in different contexts (for example which exchange rate or discount rate to use will depend on whether it is an international donor or developing country government that is making the investment decision). We have therefore provided details of calculations and present results under different assumptions so that users can tailor the estimates to their needs. More detailed discussion of the assumptions is provided in the paper below.

Criteria for Inclusion of an Evaluation in J-PAL Analyses

Studies included in the J-PAL cost-effectiveness analysis were randomized controlled trials of social programs that measured the impact of a given intervention. This excluded studies employing quasi-experimental designs such as construction of a statistically similar comparison group. Programs that were never meant to be scaled up (e.g. very high program costs that were designed to test a conceptual question not a scalable program) were not included. And to maintain a minimum standard for statistical power, only results which were significant at the 10% level or better were included.

While most studies included in the cost-effectiveness analysis have been published, we do not necessarily wait for the paper to be published in an academic journal. Once an author puts a working paper version out in the public domain, and if it meets the conditions described above, we contact the authors to confirm that the results are reportable and unlikely to change. We then construct a cost-effectiveness analysis, and if for some reason the results do change during the peer review process, we incorporate those changes.

Key Assumptions in Cost-Effectiveness Analysis:

- Analysis is done as of December 31, 2009, so GDP deflators have been used to adjust costs to 2009 prices.
- All results are reported in U.S. dollars. Nominal exchange rates have been used to convert local currencies to USD. A version with PPP exchange rates is planned for the future.
- A real discount rate of 5% has been used to discount costs and benefits.
- Transfers have been added as a cost but we will shortly be producing versions excluding the cost of transfers.
- Spillover effects have been included as long as they are carefully measured in the randomized evaluation.
- We have used estimates of average disease burdens across developing countries, rather than prevalence in the particular country to compare interventions across different regions.
- Where costs could be highly dependent on specific contexts (such as population density) we have provided ranges of cost-effectiveness based on these parameters.

Not all the adjustments described in this document apply to all the cost-effectiveness analyses conducted by J-PAL, as certain issues only arise in some studies. Some of the cost-effectiveness analyses that were completed in prior years may need to be adjusted for some of these issues, including conversion to 2009 U.S. dollars. We plan to incorporate these adjustments across all the cost-effectiveness analyses in the near future.

I. *What We Include in Calculations*

There are many different costs and benefits that can be included in a cost-effectiveness analysis. Section (a) considers transfers, or costs to the program that may not be costs to society, section (b) considers whether the multiple effects of a program should, or even could, be counted in a single cost-effectiveness analysis, section (c) considers whether spillover effects should be counted as benefits, and section (d) considers how to include the costs of items which have been provided for free in the pilot program.

a. Transfers: Costs to the Program but not to Society

Transfers, where money or goods are redistributed from one person or organization to another represent an accounting cost to the government or organization undertaking the program but not to the society as a whole. Transfers can be either cash transfers, or non-cash transfers that involve giving some goods away.

Mexico's PROGRESA program, where the government transferred cash to families conditional on their children's attendance at school and healthcare check-ups, is an example of transfers of the first type. The government's costs can be divided into administrative costs (e.g. the costs of monitoring whether children are attending school and organizing the distribution of funds) and transfer costs (the amount of money that is actually transferred to families who have complied with the conditions of the program). If we are looking at the cost-effectiveness of PROGRESA in achieving the policy goal of increased attendance at school, the benefit of the program is how many additional days of school were induced because of the promise of the cash reward.

However, the cash that is transferred to the family has additional benefits beyond simply incentivizing school attendance. For instance, it could be used to buy food and improve the nutrition of the children and the family more generally. Thus the question arises as to whether the cash transfer should be counted as a cost at all. If we define "costs" as costs to the program, then the cash transfer is a cost. If we define "costs" as costs to society then we should not include them as the value is simply transferred from one group to another—it does not get used up or diminished along the way.

If we were conducting a cost-benefit analysis (where all the benefits of a program are given a monetary value and compared to the costs), rather than a cost-effectiveness analysis, we would not face this question. The cash outlay would be a cost to the program and the cash received would be counted as a benefit and the two would cancel out. But because of the difficulties of putting a monetary value on benefits (like saving a life, additional years of schooling, or avoided discomfort from diarrhea) we have decided to only do a cost-effectiveness analysis—i.e. the cost of achieving a given policy goal.

As with many of the other methodological issues raised in this note, the "correct" treatment of transfers depends on the question being asked. If the question is "what is the most cost-effective way to reach a particular policy goal" then transfers should be treated as a cost. If the question is more broadly "is this program an effective and efficient program" then transfers should be excluded. Our approach will therefore been to show how much of the costs are administrative costs and how much are cash transfers so that policymakers can choose the approach that fits their situation.

The question of what to do with transfers becomes more complicated when we deal with non-cash transfers (giving a good or service). Take, for example, the use of lentils as an incentive for immunizing children in rural India. The purchase of the lentils represents a cost to the NGO running the program, but they also represent a value to the family that is in addition to the benefit the program has in inducing higher immunization rates (the policy goal this program seeks to achieve). If the family were going to buy lentils anyway (lentils are a major component of the diet in the area) then arguably the family is as well off receiving lentils as they would be receiving cash and the value of the transfer should be excluded when looking at cost-effectiveness for society as a whole.

However, non-cash transfers are often given because the donor puts a higher value on the good than the receiver—i.e. there is a deliberate attempt to skew the consumption basket of individuals. In this case the value of the benefit is not the same as the cost of the transfer. For example, in a program that gives free school uniforms to students in Kenya who are required to wear uniforms to attend school, we do not know if the family gets the same value from the uniform as the cost of the uniforms to the program implementers. It is possible that the family values it at less than the cost of the uniform, since some families were formerly unwilling to purchase school uniforms, but it is also possible that the program implementers value the benefit of the uniform more that its cost, since it not only provides clothing but also enables more children to attend school. Getting an accurate estimate of how much the receiving household values the non-cash transfer is hard and we have not attempted it. Hence we include the full costs of non-cash transfers in our cost analysis.[1] This will imply that the effectiveness of these programs will be understated in the cost-effectiveness analysis.

b. Multiple Aims: Effects Outside the "Scope" of the Program

Anti-poverty programs often have multiple impacts on the lives of the poor. For example, a program that gives meals at school to incentivize attendance can increase both children's attendance (measured in days of schooling) and their nutrition (measured as weight or height gain). While *cost-benefit analysis* examines the entire impact of a program by estimating the value of all of these effects, *cost-effectiveness analysis* seeks to examine the cost of causing a certain amount of change in only one indicator like days of schooling. This presents a problem, because cost-effectiveness can only examine one unit of benefit at a time (i.e. additional days of schooling *or* kilograms of weight gain), even when a program may affect all of these outcomes.

Although we cannot capture the "overall" impact of the school meals program in cost-effectiveness analysis, it is still useful to know how well this program achieves a given aim. If a policymaker's sole goal was to increase school attendance, then a high cost per additional day of schooling induced using a school meals program would be a good indicator that it was not the right program for their context. Although cost-effectiveness analysis does provide a basis for comparing the cost of achieving a specific outcome, it is an incomplete representation of all the effects of a program. We have therefore inserted the symbol § in the graphs to indicate that a program has multiple, statistically significant effects beyond what is being reported in a single cost effectiveness graph.

c. Spillover Effects: Effects on the Non-Treated Population

We have included impacts which have spilled over to non-participants in our estimation of cost-effectiveness. An example of a program with spillover effects is the provision of deworming drugs to schoolchildren to promote school attendance. Intestinal worms are spread through skin or oral contact with soil contaminated with infected fecal matter. Reducing the overall number of community members infected with worms also reduces worm transmission to children not treated with deworming drugs. Even though the program did not directly treat them, the untreated children are still more likely to attend school as a result of the overall decrease in worm infection. Therefore their additional days of schooling are incorporated in the cost-effectiveness analysis. We have only included spillover effects where they have been measured as part of a randomized evaluation.

One potential issue with including spillover benefits is that as a program is scaled up to an entire state or country, spillover effects are likely to decline. In the example of deworming, if the neighboring schools were also treated

[1] For a more complete discussion of the issues surrounding cost estimation for CCR programs, see "The Cost of Poverty Alleviation Transfer Programs: A Comparative Analysis of Three Programs in Latin America" by Caldes, Coady, and Maluccio (2004)

for worms they would not have benefited from spillovers (although out of school children would continue to benefit from spillovers). Without knowing the precise context in which a program would be implemented, however, it would be impossible to make an appropriate adjustment to the spillover impacts and we do not attempt it.

d. **Free Goods or Services**

In some evaluations, certain goods and services are provided for free to the program implementers- for example a community may donate their labor to a project or another organization may donate an input such as chlorine or vaccines. Since this is a cost to whomever is providing the input we have included the cost (or an estimate of the true cost) of such free goods or services in the cost-effectiveness analysis.

II. Issues of Currency and Time

a. Inflation and Present Value

Exchanging Currencies: Many of the evaluations report program costs in U.S. dollars (USD), but some report costs in local currencies. Before addressing issues of inflation and time discounting, all costs were uniformly converted into USD using the average annual exchange rate of the year in which they were incurred (http://www.oanda.com/currency/historical-rates). After converting the costs to USD, we applied U.S. inflation rates so that individual cost-effectiveness figures are not driven by country-specific inflation rates, thus making the resulting cost-effectiveness figures more comparable to one another. The US inflation rate also has the advantage of being relatively stable, so programs piloted in years with radically different local inflation rates will not be driven by those episodes of abnormally high or low inflation.

Impact of Inflation: When costs are incurred over a number of years and reported in terms of their nominal amounts in the years in which they were incurred, we converted the costs to their equivalent value in constant terms of the base year (the year when the program began). We have used the average GDP deflators during the duration of the program, rather than consumer price indices as the measures of inflation since they cover a wider range of goods and services of the kind used in most anti-poverty programs. When the year the program began (base year) precedes the year of our cost-effectiveness analysis (2009), we have adjusted the total program cost (the value of the real cost stream as discussed below) so that it is expressed in terms of 2009 amounts, using the average GDP deflators, taken from the World Bank Data Finder (http://datafinder.worldbank.org/inflation-gdp-deflator).

Time Value of Money: When both costs and benefits were accrued over many years, we discounted both the costs and the benefits back to the starting year of the program (present value) using a 5% discount rate. There is no universally applicable real discount rate in the literature.[2] The "right" discount rate depends on who is making the investment: different funders will have different social time preference rates and costs of capital. We use a 5% real rate to reflect a reasonable measure of social time preference and the long-run real cost of capital for a range of governments (the richest countries will have lower costs than this, and the poorest higher). This gives us the total program cost in terms of base year prices.

[2] There are signicant variations in public discount rate policies practised by countries around the world, with developing countries in general applying higher social discount rates (8–15%) than developed countries (3-7%). http://www.adb.org/Documents/ERD/Working_Papers/WP094.pdf

Example 1: PROGRESA

A program of conditional cash transfers was piloted between 2000 and 2001, and we have actual cost data from that program's implementation. In 2000, implementers spent $1,149,000, in 2001 they spent $1,348,000, and in 2002 they spent $1,492,000. We are performing our analysis in the year 2009.

1. We begin by making sure all costs are expressed in USD of the year in which they were incurred. If costs were given to us in the local currency, we would begin by converting them to USD using the average annual exchange rate of the year in which costs were reported. This step is not necessary in this example.

2. Because we have actual cost data reported in nominal terms of the year in which they were incurred, it is necessary to bring all cost figures into real terms of the base year of the program, in terms of the local currency.

Cost	2000	2001	2002
Program Costs (USD in Year Incurred)	$ 1,149,000	$ 1,348,000	$ 1,492,000
Average US Inflation, 2000-2002	2.1%	2.1%	2.1%
Elapsed Years	0	1	2
Inflation Adjusted Costs (2000 USD)	$ 1,149,000	$ 1,320,274	$ 1,431,256

3. We can then estimate the present value of this cost stream in the base year.

Cost	2000	2001	2002
Inflation Adjusted Costs (2000 USD)	$ 1,149,000	$ 1,320,274	$ 1,431,256
PV in Start Year of Program (2000 USD)	$ 1,149,000	$ 1,257,404	$ 1,298,191
PV of Cost Stream (Sum, in 2000 USD)	$ 3,704,595		

4. The present value of the cost stream expressed in terms of 2000 USD is $3,704,595. But since we are performing the cost-effectiveness analysis as in 2009, we want to estimate what this cost stream would be on that date. This is done by applying the average US GDP inflation rate for 9 years (approximately 2.5%) as follows:

Cost	Base Year
PV of Cost Stream (2000 USD)	$ 3,704,595
Average US Inflation, 2000-2009	2.5%
Elapsed Years	9
PV of Cost Stream (2009 USD)	$ 4,626,532

The present value of the cost stream the beginning (base) year of the program, if the program were to begin in 2009, is $4,626,532.

Example 2: Chlorine Dispensers

A program to install chlorine dispensers was piloted in 2009, the same year in which we are performing our analysis. This program will cost $1000 to install the dispensers, plus an estimated $50 each year in maintenance for the next 15 years totaling $750. Note that the maintenance cost is expressed in terms of 2009 dollars; the nominal cost of maintenance in subsequent years will likely be higher.

1. The first step, bringing all costs in terms of base year dollars, is not necessary because the cost stream is already expressed in terms of 2009 dollars.

2. It is necessary to estimate the present value of the cost stream, since we have costs in multiple years. This is relatively simple in this example because the project will be incurring the same cost every year, and so we can use the Present Value formula.

Costs	2009
Annual Maintenance Cost, in 2009 USD	$ 50
Discount Rate	5%
Year Since Costs Counted	15
Time Discounted Costs	$ 519

3. In this case it is not necessary to bring the costs in terms of the year of analysis, because the base year of the program and the year of analysis are the same.

So our final estimate for the present value of the cost stream in the start year of the program, if the program began in 2009, is $519.

b. Exchange Rates & Purchasing Power Parity

If the cost effectiveness estimate resulting from our inflation and present value adjustments is in terms of local currency, we converted it into U.S. dollars using nominal exchange rates. In the future, we plan to create a version using purchasing power parity (PPP). For an international donor considering whether to invest in a deworming program in Kenya or a school feeding program in Brazil, a comparison using nominal exchange rates is most useful when the cost numbers also come from evaluations that were done in those two respective countries.

However, if a policy maker in (say) India wants to compare the cost-effectiveness of deworming with the cost-effectiveness of school feeding in India, and we have data on deworming from an evaluation in Kenya and data on school feeding from an evaluation done in Brazil, purchasing power parity is a better measure as it gets closer to putting the two programs on a similar cost basis (even though it does not fully adjust for the fact that similar programs will cost different amounts in different countries). In the future therefore, we will also provide PPP adjusted cost effectiveness analysis.

III. Estimating Impacts

There are several issues which have informed our decisions on which impacts to include in our cost-effectiveness analysis and how to calculate them.

a. Significance of Effects

We have only counted the impacts of effects that were significant at the 10% level or better. If an intervention was found to have zero impact (with at least 90% confidence), then we have included such zero impacts in the analysis, too.

b. Proximal Success vs. Final Impact of Programs

The aim of impact analyses is to show not just the relative costs of different channels of distributing goods or services to the poor, but how those goods and services translate into impacts, and what the impact for a given expenditure is. We therefore make a distinction between the proximal "success" of a program (how well it achieves its proximal goal), and the final "impact" which is a result of that proximal success. For example, studies that promote chlorination of water to prevent diarrheal disease may measure success in terms of number of additional houses with chlorinated water (proximal success), but the objective of the program is to reduce diarrheal disease (final impact), rather than to simply get chlorine into water. Wherever possible (i.e. where there is the necessary evidence from other studies) we translated proximal effects into measures of their final impact, to give a better sense of the final outcomes of a program in terms of its original goals. In doing this, we used generally agreed upon estimates of the way in which proximal effects translated into final impacts based on rigorous studies.

c. Distance Along the Marginal Cost/Benefit Curve

An intervention may have different impacts in different contexts due to differing starting levels of the underlying problem, and this means that different countries or regions may be at different points on the marginal cost/benefit curve. This can result in variations in the cost-effectiveness figures for programs piloted in different regions. For instance, a chlorine tablet should be equally effective at killing bacteria in Africa or in Asia, but the amount of diarrheal disease that is prevented as a result of improved water quality depends on the pre-existing levels of diarrheal disease in those two places. Even though chlorine can (if used perfectly) reduce bacteria to nearly zero in both cases, the elimination of diarrhea-causing bacteria may prevent five diarrhea incidents per child annually in Africa, but only two in Asia, because of lower starting quality of water and higher incidence of diarrhea in Africa.

To mitigate this problem, we have used several approaches- tailored to the particular situation. In some cases, especially among education interventions, we have measured impacts in terms of standard deviations. For other programs, rather than using the country-specific estimates of underlying problem, we have used global estimates from meta-analysis. For example, in the cost effectiveness analysis of various interventions in Kenya and Pakistan to reduce diarrhea, rather than using the starting number of diarrhea incidences per child in Kenya and Pakistan, we use the average number of diarrhea incidences per child across developing countries as the starting level of diarrhea incidence.

IV. *General Limitations of Randomized Evaluations and Cost-Effectiveness Analysis*

While cost-effectiveness analyses are intended to provide a basis for comparison between programs with similar objectives, there are some limitations to cost effectiveness calculations and to the drawing of conclusions about individual programs.

a. Partial vs. General Equilibrium

Randomized evaluations provide a snapshot view of what the partial equilibrium effects of a program will be. Some randomized evaluations are designed to pick up the long-term impacts of a program but many pick up shorter run effects. Even when an evaluation is designed to pick up long- run effects for a particular cohort, it is possible that people will change their response to a program as it becomes more established—i.e. later cohorts may respond differently—or the benefits of the program will change as the program is scaled up. For example, graduates of a vocational education program in a rural area could be expected to see very high returns at the outset, since they are the only ones with specialized knowledge (see also the earlier section on Distance Along the Marginal Benefit Curve). But as time goes by and more people graduate from the vocational education program, the supply of educated workers would increase and the returns to vocational education could decrease as a result of excess supply.

Spillovers may also decline as programs become universal. On the other hand, benefits sometimes become larger as programs are scaled up as behavior change is reinforced by seeing others undergo the same behavior change. The marginal benefits of education can also increase as more people become educated if there are complementarities between skilled workers. It is difficult in general to precisely estimate the extent to which general equilibrium impacts may be different from partial ones, although individual studies often discuss the issues in a particular context.

b. Experimental and Scalable Modes of Implementation

The costs of a program evaluated in its pilot phase may be different from the actual costs if one were to massively scale-up the program. This is because there may be advantages to working on a larger scale, such as purchasing supplies in bulk, which could increase a program's cost-effectiveness. But there may also disadvantages, such as the increased difficulty of administering a program over a wide area, which could decrease cost-effectiveness. Moreover, some components of the program may be altered when it transitions from a pilot evaluation to a large scale-up.

For this reasons, where changes have been made to the original program design in the scaled-up version, it may be useful to conduct an evaluation of a pilot to verify the program's impact. Similarly, if there is reason to believe that the costs of the scale-up are likely to be different from the original evaluation, it may be useful to perform a detailed survey of local costs before expanding the program. As more pilot evaluations are scaled-up in the next few years, there will be a better understanding among researchers and policy makers about how costs in an evaluation translate into costs in a large scale-up.

c. Cost Sensitivity to Various Parameters

In cost-effectiveness analysis, it is necessary to incorporate some features of individual programs as assumptions about the general implementation of the program. While normalizing the assumed pre-existing levels of disease, etc. can help to ensure the comparability of impact figures (see the earlier section on Distance Along the Marginal Benefit Curve), there are other location-specific parameters that have an effect on the cost-effectiveness of a program. For example, in the area of Kenya where chlorine dispensers were tested, an average 31 families used each water source (spring) where the dispensers were placed. In a more densely settled area with more families per spring, the cost-effectiveness of such an intervention would improve because more families would receive the benefit of chlorine, while the fixed cost of installation and maintenance would stay the same.

In addition to complicating comparison, the sensitivity of cost-effectiveness numbers to certain parameters can also create the misleading impression that if a program were to be implemented in *any* area, the cost per impact would be the same as in our cost-effectiveness example. While our numbers are intended to provide a means of comparison between different programs, they are not intended to reflect exactly what the program would cost to implement in any setting. To reflect the way that costs per impact can be expected to vary with certain parameters, we have chosen the most relevant variable for a given program and shown how the cost-effectiveness of that program would change with that variable over a reasonable range. Using the example of chlorine dispensers, we have provided a point estimate of cost-effectiveness at 31 families per spring, as well as how many diarrhea incidents could be avoided if there were 16 and 46 families per spring.

More detailed analysis is, in most cases, available to help those wishing to scale up and implement programs for which cost-effectiveness results are given in different contexts. This detailed analysis can be tailored to include revised cost estimates taken from local knowledge of the costs in specific contexts and country specific prevalence rates, and generally adapt the figures shown in our analyses to the needs of different users.

d. Contextual Differences

Though we have done our best to ensure comparability between the cost-effectiveness figures, there will always be contextual differences that will affect the costs and impacts of program implementation. These can include:

- Differences in the quality or level of existing infrastructure
- The efficiency of the local group implementing the program
- Differences in local costs due to traditions of community contribution in the form of labor or donations
- Price differences (fact that the law of one price doesn't hold, even when currency exchanged)
- Local predispositions for or against certain products
- Different motivations among the beneficiaries

Conclusion

Despite the many caveats and judgment calls described above, the cost-effectiveness calculations presented here are in our view an important aid to policy making. Often the differences in cost-effectiveness of alternative programs are sufficiently large that small changes in assumptions do not change the underlying conclusions very much, especially if these assumptions are uniformly applied across all programs. In other cases, thinking through why the results are what they are and what would be needed to make an alternative option more cost-effective is an exercise that can be very helpful to policymakers as they make the decisions about the best way to achieve their policy goals with limited resources.