Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries:

A General Framework with Applications for Education

Iqbal Dhaliwal, Esther Duflo, Rachel Glennerster, Caitlin Tulloch[1]

Abdul Latif Jameel Poverty Action Lab (J-PAL), MIT

December 3, 2012

Abstract

In this paper we discuss how comparative cost-effectiveness analyses can help inform policy in developing countries and the underlying methodological assumptions necessary for performing this kind of analysis. This paper does not suggest a single set of "correct" assumptions, because the assumptions adopted in a cost-effectiveness analysis should reflect the perspective of the intended user. Rather, we discuss the issues surrounding many of these assumptions, such as what discount rate to use or whether to include cash transfers as program costs, and make recommendations on which assumptions might be reasonable given the perspective of a policymaker allocating resources between different projects. Examples are drawn from the education field to illustrate key issues and focus on specific applications to education. We hope this paper will contribute to the development of a more standard methodology for cost-effectiveness analyses and a better understanding of how these analyses can be created and used.

[1] We are grateful to Abhijit Banerjee, Paul Glewwe, Jere Behrman, Miguel Urquiola, and Patrick McEwan for valuable discussion and feedback. We also thank participants at the Minnesota conference on "Education Policy in Developing Countries" for their comments and many colleagues at J-PAL including Mary Ann Bates, Cristobal Marshall, Leah Horgan, Dina Grossman, Anna Yalouris, and Shawn Powers.

1 Introduction to Cost-Effectiveness Analysis

In the last fifteen years there has been a sharp increase in the number of rigorous evaluations of the impact of development programs in a host of fields including education, health, environment, agriculture, finance, and governance. One of the major objectives of such studies is to provide evidence to policymakers on what works and does not work in the fight against poverty, so they can use scientific evidence to determine which policies and programs to adopt and invest in.[2] But it can be very difficult for policymakers to compare results from different programs and their evaluations, performed in different countries, in different years, and that use different instruments to achieve the same outcome. For instance, studies have evaluated the impact on years of schooling of deworming programs in Kenya, conditional cash transfers in Mexico, providing free uniforms in Kenya, and providing information to parents in Madagascar. Faced with a growing body of evidence from field research and given their time and resource constraints, policymakers can find it very hard to analyze and interpret the results of multiple studies, most of which are published in technical or academic journals. As a result, policymakers may decide to ignore such evidence altogether and go back to relying on their instincts on what works or does not work, or selectively choose studies that support their instincts or predetermined choices.

One way to encourage policymakers to use the scientific evidence from these rigorous evaluations in their decision making is to present evidence in the form of a cost-effectiveness analysis, which compares the impacts and costs of various programs run in different countries

[2] We use the term "policymakers" to refer not only to civil servants in governments, but also to decision makers in foundations, international development organizations and non-governmental organizations (NGOs) who make decisions regularly on how to allocate resources between competing programs that try to achieve the same objective.

and years that aimed at achieving the same objective. Some earlier work has attempted to compare the relative cost-effectiveness of different education programs within a particular context. For example, Kremer, Miguel, and Thornton (2009) show the relative cost-effectiveness of different programs carried out by International Child Support in Kenya; while Banerjee, Cole, Duflo, and Linden (2007) compare education programs run by the NGO Pratham in India. Limiting an analysis to programs carried out by the same organization in the same country makes it easier to ensure that costs and impacts are calculated using the same methodology, but it restricts the range of interventions that can be compared. Drawing comparisons across projects performed in different countries, by different organizations, and in different years, as we discuss in this chapter, raises many more questions about how to ensure comparability, but it also holds the promise of being a more useful tool for policymakers. Examples exist of this kind of analysis, such as Carnoy (1975), although they seem to be infrequent in the literature. A major challenge in this kind of analysis is to strike the right balance in the trade-off between the need for policymakers to see comparisons in a form that is intuitive and easy to understand and the need to present enough information to help them appreciate the finer nuances of the programs, including the programs' sensitivity to various factors such as population density or certain large input costs. We believe that the value of promoting the use of scientific evidence in policy making is sufficiently high that it is valuable for researchers to create such analyses, while explicitly stating their assumptions and clearly acknowledging the limitations.

Cost-effectiveness analysis by itself does not provide enough information for a policymaker to make an investment decision, but such analysis does provide a very useful starting point for researchers and policymakers to collaborate in assessing the efficacy of the different programs and their relevance to the particular situation. Cost-effectiveness analysis results, with detailed information on underlying costs and impacts, combined with an

understanding of the problem being addressed and of other contextual factors such as current input prices and local institutions, can provide important insights into which programs are likely to provide the greatest value for money in a particular situation, and to identify the key factors to which these outcomes are most sensitive. When cost-effectiveness analyses have been done with data at a highly disaggregated level, where assumptions about key factors such as program take-up or unit costs are made explicit, it is much easier to perform sensitivity analysis. This sort of sensitivity analysis gives policymakers an idea of how cost-effective a similar program might be in their situation by varying key assumptions to reflect their context.

There is a substantial literature on how to conduct cost-effectiveness and cost-benefit analyses, much of which has been written for the assessment of domestic policies in the US or other developed countries (Levin and McEwan 2001; US Department of Health and Human Services 1996). But there is often no single right methodology—the appropriate assumptions to make usually depend on the precise question being asked or how the analysis will be used. For example, whether to include user costs as a cost of the program will depend on whether the objective of the policymaker or implementer is to maximize cost-effectiveness of the implementer or of society as a whole. This chapter examines many of these questions by presenting a standardized approach to applying cost-effectiveness analysis to inform educational policymaking in developing countries.

In this chapter we discuss various alternative assumptions and methodologies, which of these are most appropriate in what situation, and why J-PAL as an organization has chosen the particular approach that we use in constructing comparative cost-effectiveness analyses. We illustrate our discussion of cost-effectiveness methodology with an analysis of programs that seek to increase student attendance in different countries. Because we recognize that different

policymakers may have different perspectives, we also show how sensitive these results are to different assumptions or approaches, and throughout our analysis we place an emphasis on transparency so that users can understand the various components of the analysis and how it should, and should not, be interpreted. We also suggest some ways to illustrate sensitivities in a way that can be reasonably easy to interpret and include examples of user-friendly sensitivity data throughout the chapter, as well as in Appendix A. However, it is impossible to include every alternative way of showing the results in one chart, and eventually decisions based on trade-offs between various alternatives have to be made so that a useful picture of relative cost-effectiveness emerges.

Such cost-effectiveness analysis requires detailed underlying cost and impact data. Currently, most published articles evaluating social sector programs in developing countries, including education, do not provide enough specific cost data to undertake a good cost-effectiveness analysis (Levin 2001). For an excellent listing of the existing articles that do focus on educational cost-effectiveness analyses in the developing world, see the bibliography of Levin and McEwan's book "*Cost-Effectiveness Analysis*." One of our objectives moving forward is to encourage researchers to record detailed cost and impact data (ideally on a standardized basis) and make underlying calculations publicly available so that more, and more complex and rigorous, cost-effectiveness analyses can be done in the future.

1.1. Goals of Cost-Effectiveness Analysis. Cost-effectiveness analysis, in the simplest terms, calculates the ratio of the amount of "effect" a program achieves for a given amount of cost incurred, or conversely the amount of cost required to achieve a given impact. For program evaluation, this means measuring the impact of a program in achieving a given policy goal (for example, the extra years of schooling induced) against the cost of the program. This ratio, when

calculated for a range of alternative programs addressing the same policy goal, conveys the relative impacts and costs of these programs in an easily understandable and intuitive way. However, relatively few studies published in academic journals include cost data on the programs they are evaluating, and what data are available are presented in a wide variety of formats that does not allow for easy comparison between programs. Moreover, what exactly is meant by "costs" and "impacts" is itself subject to considerable debate, depending on the perspective from which the analysis is being undertaken. Are the costs to all stakeholders relevant, or only those that accrue to the implementing organization? Can multiple effects on a number of outcomes be included in the measure of "effectiveness?" To think about these questions, it is important to first explicitly state the goals of cost-effectiveness analysis.

The value of cost-effectiveness analysis is twofold: first, its ability to summarize a complex program in terms of an illustrative ratio of effects to costs, and second, the ability to use this common measure to compare multiple programs evaluated in different contexts and in different years. The first requires technical correctness with respect to the program's actual costs and impacts as they were evaluated, while the second requires adherence to a common methodology for estimating costs and effects across various studies. For cost-effectiveness analysis to be useful and informative, it must maximize the comparability of estimates for different programs without straying from a correct and complete representation of the costs and effects of each program as it was actually evaluated. When done correctly, such analysis can be a useful tool for decision makers in organizations that fund or implement education and other social programs in developing countries, allowing them to compare the results of alternative programs when deciding how to allocate resources. This includes funders (such as foundations and international development organizations), and governments and NGOs that both fund and implement programs.

1.2. Why Cost-Effectiveness Analysis Rather than Cost-Benefit Analysis? Cost-effectiveness analysis shows the amount of "effect" a program achieves on one outcome measure for a given cost, while cost-benefit analysis combines all the different benefits of a program onto one scale (usually a monetary scale) and shows the ratio of the combined benefits to cost. The advantage of cost-benefit analysis is that it makes it easier to assess a program with multiple outcomes. Additionally, putting both costs and benefits onto the same scale delivers not just a relative but an absolute judgment: whether or not a program is worth the investment, and which program among several yields the best rate of return. A good example of where cost-benefit analysis is most useful is a program that involves an up-front investment (say, the building of a new hospital) that will generate a stream of benefits (e.g. reduced maintenance costs) in the future. Apply the cost of capital as the discount rate and the result will tell you whether the investment is worthwhile.

The downside of using cost-benefit analysis is that it requires a number of assumptions about the monetary value of benefits on which different organizations may have very different views. When an organization's value of statistical life or years of education is known, then cost-benefit gives very concrete answers. But from a general perspective, where readers may place very different values on outcome measures, a single cost-benefit analysis may not be generally applicable. In the calculation of disability adjusted life years (DALYs), for example, there is disagreement about whether to give different weights to the health of people of different ages (Anand and Hanson 1997), and even once the number of DALYs averted has been calculated for a program, there is no standard monetary value per DALY. Cost-effectiveness analysis allows for the user to apply their own judgment about the value of the benefits. The analysis tells the user what can be achieved for what cost and leaves it to the user to decide whether that benefit is worth the cost. In the case of education in developing countries, cost-benefit analysis would

require, among other things, estimating the increase in productivity achieved as a result of an increase in school quality or quantity. A monetary valuation of any improvement in health, intergenerational benefits, and the pure consumption benefits of education could also be included. However, there are few commonly agreed upon monetary values for outcomes such as years of life or increases in test scores, making it difficult to create a single cost-benefit analysis that would be useful for a wide range of organizations.

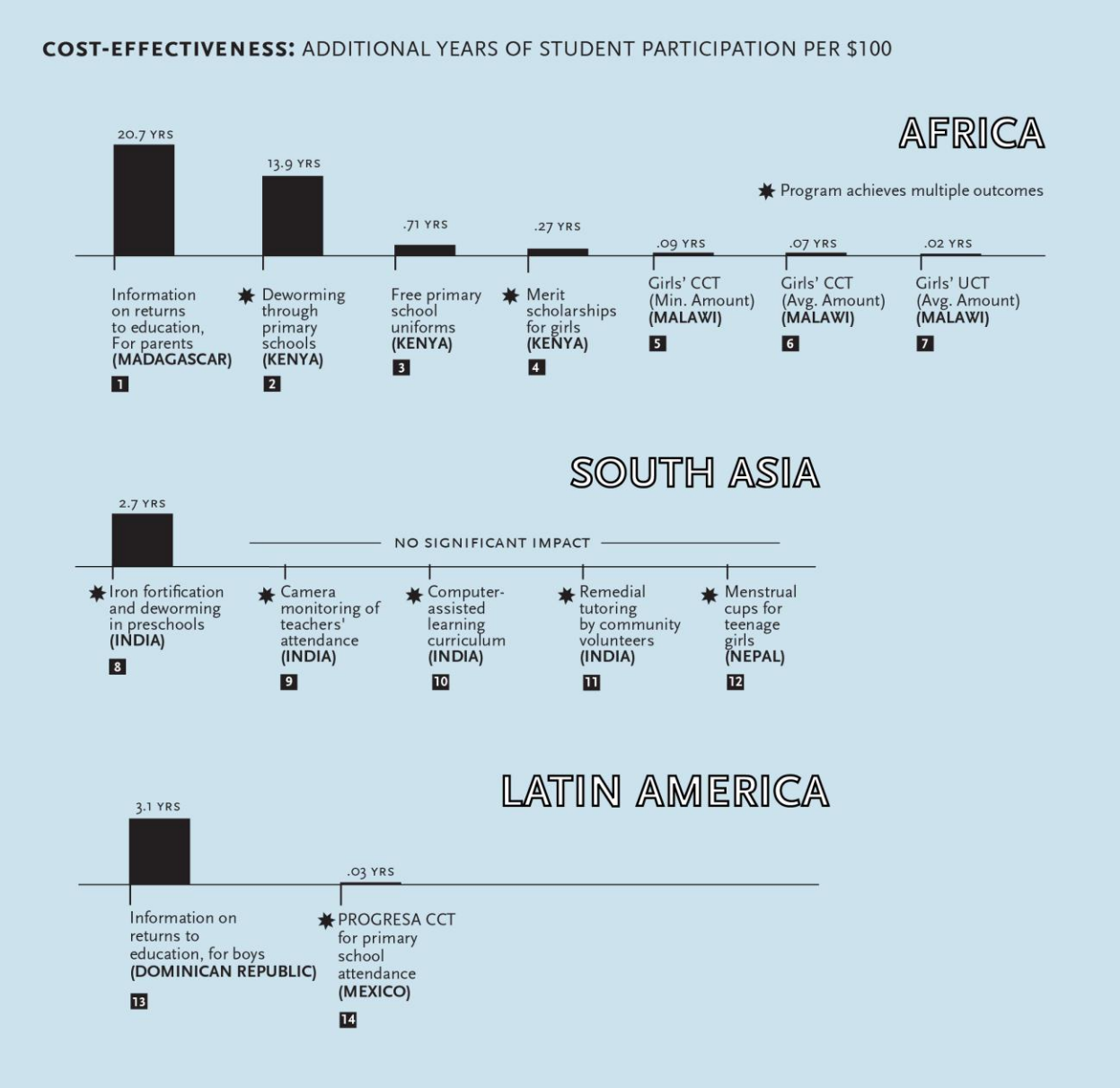
1.3. Cost-Effectiveness Analysis in Education. For cost-effectiveness analysis to be a useful alternative to cost-benefit analysis, however, it is necessary to agree on an outcome measure which would be the "key objective" of many different programs and policymakers. In the field of education there are a few obvious contenders. Two of the Millennium Development Goals focus on children's school enrollment. Although time spent in school is an imperfect measure of the increase in education, it does provide a useful approximation, particularly given the recent focus on increasing primary school enrollment and attendance. Similarly, there is increased attention on the need to reduce provider absenteeism, and standard methods of measuring teacher absenteeism are emerging. J-PAL recognizes that these are both important aspects of improving education and is undertaking cost-effectiveness analyses of both of these outcome measures. For the purposes of this chapter, we will use the cost-effectiveness analysis for student attendance and enrollment to illustrate key issues. This analysis includes eleven programs from six countries, all of which were assessed using a randomized evaluation[3].

[3] These programs were analyzed from the perspective of 2010 in USD. The discount rate used was 10 percent, all exchange rates were standard, and inflation was calculated using GDP deflators.

Table 1. Programs Included In Cost-Effectiveness Analysis of Student Attendance and Enrollment

	Project	Country	Researchers	Publication	Base Year
1	Information Session on Returns to Education, for Parents	Madagascar	Trang Nguyen	"Information, Role Models, and Perceived Returns to Education: Evidence from Madagascar"	2006
2	Deworming Through Primary Schools	Kenya	Edward Miguel, Michael Kremer	"Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities"	1999
3	Free Primary School Uniforms	Kenya	David Evans, Michael Kremer, Mũthoni Ngatia	"The Impact of Distributing School Uniforms on Children's Education in Kenya"	2002
4	Free School Uniforms	Kenya	Michael Kremer, Edward Miguel, Rebecca Thornton	"Incentives to Learn"	2001
5	Iron Fortification & Deworming in Preschools	India	Gustavo Bobonis, Edward Miguel, Charu Puri-Sharma	"Anemia and School Participation"	2002
6	Camera Monitoring of Teachers' Attendance	India	Esther Duflo, Rema Hanna, Stephen Ryan	"Incentives Work: Getting Teachers to Come to School"	2003
7	Computer-Assisted Learning Curriculum	India	Abhijit Banerjee, Shawn Cole, Esther Duflo, Leigh Linden	"Remedying Education: Evidence from Two Randomized Experiments in India"	2001
8	Remedial Tutoring by Community Volunteers	India	Abhijit Banerjee, Shawn Cole, Esther Duflo, Leigh Linden	"Remedying Education: Evidence from Two Randomized Experiments in India"	2001
9	Menstrual Cups for Teenage Girls	Nepal	Emily Oster, Rebecca Thornton	Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation	2006
10	Information Session on Return to Education, for Boys	Dominican Republic	Robert Jensen	"The (Perceived) Returns to Education and the Demand for Schooling"	2001
11	Cash Transfers Conditional on Attendance at Primary School	Mexico	David Coady, T. Paul Schultz	"The Application of Social Cost-Benefit Analysis to the Evaluation of PROGRESA" "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program" "Final Report: The Impact of PROGRESA on School Enrollment"	1997

The graph showing this analysis is in Figure 1.



The analysis shows that the cost-effectiveness of programs to increase children's time in school varies widely, with informational campaigns in the Dominican Republic and Madagascar and school-based deworming in Kenya providing the greatest value for money among the programs examined here. Both of these programs can be delivered inexpensively and cause large increases in student attendance and enrollment. Programs that reduce the costs of schooling through subsidies or provide incentives conditional on attendance also increase time in school, but at a higher cost, and conditional cash transfer programs are not as cost-effective as a program to

increase time in school. As we discuss below, it is important to think about context and the sensitivity of these results to other factors when drawing conclusions for new programs based on these numbers.

A more challenging question for cost-effectiveness analysis is how to appropriately measure the quality of learning in a comparable way across studies. Some educationists believe that test score measures fail to capture some important aspects of learning. We will not engage in that discussion here. Instead we are concerned with how to appropriately compare gains in test scores in one context with such gains in another context. For instance, how do you compare a seven-year-old boy in India learning to recognize letters with a thirteen-year-old girl in Colombia learning the chemical composition of water? There are internationally standardized tests available that could be used, such as the Program for International Student Assessment Test (PISA), but these are often at too advanced a level to detect changes in test scores in poor countries. The NGO Pratham's rapid assessment tests are a useful tool for testing literacy and basic math skills across countries and have been widely used in India, Pakistan, Tanzania, Kenya, and Morocco. Most education programs, however, are affecting learning in between these two extremes, and the majority of education evaluations in developing countries therefore use tests that are tailored to the specific context to measure learning outcomes. One practical approach is to use the standard deviation of scores in the control group as the scale against which impact is measured, as is quite common in the education literature.[4] This is the approach that J-PAL is taking in measuring the cost-effectiveness of programs aimed at increasing learning. The results of this analysis are not presented here, as it is still ongoing.

[4] See, for example, Glewwe, Ilias, and Kremer 2010, Burde and Linden 2010, Nguyen 2008b, and Linden 2008.

1.4. Defining the Perspective of Users of Cost-Effectiveness Analyses. As discussed above, the appropriate methodology to use when doing a cost-effectiveness analysis usually depends on the perspective of the policymaker who will use it. The methodology that is adopted in the examples given here (and by J-PAL as an organization) is intended for an audience of policymakers in governments, foundations, international development agencies, and NGOs that have a particular policy objective in mind and are trying to decide between a range of different options for achieving that policy objective. We are not trying to help the prime minister of a country, or the chairman of a foundation, to decide whether to put their money into education versus health. In our view, we do not have enough information to help them make that decision, which should reflect the specific social preferences in that country or the mission of that foundation. Instead we are taking the perspective of, for example, the minister of education of a state in India, or an education program officer in a foundation, who aims to maximize their impact on a particular objective such as student attendance within a budget constraint. We assume that the policymaker cares not just about their own budgetary costs, but also about the costs that a particular program will impose on the beneficiaries—they are presumably involved in these decisions because they wish to help the beneficiaries of their programs. This perspective influences a number of judgments we make in this chapter.

1.5. The Challenge of "Comparative" Analyses. Because cost-effectiveness analysis is intended as "an input into resource allocation decisions concerning a wide spectrum of alternative programs" (U.S. Department of Health and Human Services 1996), it is necessary to provide comparable figures for programs whose costs and impacts were accrued in different countries, years, and institutional contexts. For example, programs may have been paid for in different currencies (e.g. 2008 dollars versus 1999 pesos) and evaluated with slightly different outcome measures (i.e. percentage change in student attendance versus number of days of

schooling gained). But cost-effectiveness analysis requires that these units be harmonized so that the cost-effectiveness ratios for all programs in an analysis are expressed in the same units. There are two primary challenges in arriving at comparable estimates: applying a common methodology to varying sets of data, and making appropriate adjustments to reflect different time, currency, and inflation rates. The rest of the chapter addresses these two issues and is organized as follows: Section 2 discusses methods of quantifying program impacts in a standard manner, including spillover and secondary effects. Section 3 reviews which costs should be included based on the perspective outlined above and how to assemble accurate cost data from available resources. Section 4 discusses the standardization of both costs and benefits into "standard units," accounting for inflation, exchange rates, and varying streams of costs and benefits. Finally, Section 5 reviews some more general issues with cost-effectiveness analysis, including the generalizability of costs and effects and a discussion of partial and long-term equilibrium effects.

2 Quantifying Impacts

In this section we discuss a number of issues related to the calculation of impact.

2.1. Sources of Impact Estimates. Many reports and studies attempt to assess the impact of education programs in developing countries. These range in quality from the anecdotal to the highly rigorous. In the examples presented here, we have chosen to include only randomized evaluations. This is not because we think that only randomized evaluations are rigorous or that there is nothing to be learned from non-quantitative studies. However, this cost-effectiveness work has been undertaken under the auspices of J-PAL, which applies randomized evaluations to social programs to understand what works or does not work in the fight against

poverty, and randomization provides a transparent criterion for selection of studies. As a result we have excluded some rigorous non-randomized studies (including some done by the authors), but we believe that this is compensated for in the resulting transparency of the selection process. By being transparent in our methodology, we make it possible for others to add more programs to our comparisons.

2.2. Programs Achieving Multiple Impacts. Cost-effectiveness analysis, by definition, focuses on the ratio of costs a program incurs to progress it causes in *one* outcome measure, but anti-poverty programs often have multiple impacts on the lives of the poor (see, for example, Chapter 5). This means that, in some cases, the chosen outcome measure may not reflect the full set of impacts of the program being analyzed. Giving children free school meals increases attendance at preschools, and its cost-effectiveness can be expressed in terms of cost per additional years of attendance. But school meals may also improve children's nutritional status, an additional impact that is "bought" with the same dollars as the increased attendance (Kremer and Vermeersch 2004).

This is an issue which cannot be easily resolved within the framework of cost-effectiveness analysis, which deliberately focuses on a single outcome measure of compelling interest to policymakers. As discussed above, cost-benefit analysis may be more suited to comparing programs with multiple outcomes, although it does so at a cost of reduced transparency. In some cases it may be possible to separate out the proportions of a program's costs that are "responsible" for different impacts. An example might be a conditional cash transfer program that offers incentives for both school attendance and regular doctor appointments for children. If the incentives are given separately, the impacts on education could be assumed to be mainly from the education subsidy rather than the health subsidy (unless one

believed that improved health contributed significantly to increased school attendance). In this case it might be appropriate to include only the costs of the education subsidy in the cost-effectiveness analysis for education, and only the costs of the health incentive in the cost-effectiveness analysis for health. On the other hand, separating out the costs of overhead and administration can be much harder.

An alternative approach, which some researchers are experimenting with, does not attempt to allocate costs by outcome, but accepts that a program is a package and should be assessed as such. In this case, if it is possible to allocate total costs between outcome measures in such a way that the effect-to-cost ratio for all the different outcome measures is superior to the best alternative method of reaching the outcome, the program is clearly cost-effective. Take deworming as an example: deworming achieves both health and education outcomes (Miguel and Kremer 2004, see also Chapter 4). If we split the costs of deworming and allocate half to student attendance and half to child health, it would be possible to calculate the cost per additional year of schooling and the cost per DALY saved. As both of these figures would indicate a highly cost-effective program for the outcome in question, we could conclude that the program was cost-effective as a package.

As is usually the case, the appropriate methodology depends on the perspective of the user of the cost-benefit information, or the precise question being asked. For example, an education minister with a fixed budget and an objective of reaching the Millennium Development Goal of universal primary education would want to know the cost-effectiveness of deworming with all the costs allocated against the student attendance objective. Similarly, the most relevant analysis for the Global Alliance for Vaccines Initiative (GAVI), which has an objective of increasing coverage of childhood immunizations, would be the full cost of various

programs against the single outcome measure of increased immunization. Instituting a conditional cash transfer program that pays beneficiaries for regular visits to health clinics would be a relatively expensive way for GAVI to achieve their objective. In contrast, the Mexican government, which has a multiplicity of objectives including health, education, and income redistribution, would probably not want to make a judgment on whether to continue with a conditional cash transfer program based on a cost-effectiveness analysis of PROGRESA that considers only a single outcome. Where we have evidence that a program achieves multiple outcomes beyond the one addressed in the cost-effectiveness estimation, we flag that study as "achieving multiple outcomes" on the cost-effectiveness graph. And if it is possible to clearly separate program costs between multiple outcomes as discussed above, we attempt to do so.

2.3. Imprecision in the Estimation of Impact. Estimates of impact can be measured only with a limited level of precision. Depending on the power of the underlying evaluation, different impact estimates will be measured more or less precisely. The point estimate of impact is typically used to calculate cost-effectiveness, at least where the impact is found to be significantly different from zero. However, comparative analyses are concerned with the *relative* cost-effectiveness of different programs, and it is quite possible that while one program may appear more cost-effective than another using point estimates of impact, the two may not be significantly different from each other if the variance around the two point estimates is taken into account.

The first question this raises is what level of significance is an acceptable criterion for the program to be included in the cost-effectiveness analysis. Programs whose impact is significant at 10% or better are included in J-PAL analyses. Having chosen a cutoff level of significance, there is a further question about what to do with insignificant impacts. Insignificance could

represent one of two things: an estimate that is quite precisely measured and is close to zero, or one that is very imprecisely measured and where a moderate or large impact cannot be ruled out. For these two kinds of insignificant results, it is not immediately clear whether to include such programs in an analysis, and what point estimate of impact, if any, to use in calculations of their cost-effectiveness.

If a point estimate is insignificantly different from zero and precisely estimated, we often say that the estimated impact is zero ("the program did not work"), even though it is rarely the case that the point estimate is exactly zero. We believe that it is important to disseminate information about which programs *do not* work, as well as those that do; thus, we have chosen to include studies that show precisely estimated insignificant impacts on our graphs. However, rather than showing a bar calculated with the insignificant point estimate, we include a space for the program on the graph with a label indicating that the program had no *significant* impact. In addition to technical clarity, there are practical reasons for labeling these programs as "no significant impact," rather than including cost-effectiveness estimates for them. This approach avoids including cost-effectiveness calculations with insignificant negative impacts or displaying very large bars in the chart if cost-effectiveness is calculated as cost per unit of impact.

It is more complicated to consider results which are not statistically different from zero, but which have wide confidence intervals including large positive or negative values for the point estimates. Because they are imprecisely estimated, we have less evidence as to the actual impact of the program. For this reason, we have chosen to exclude imprecisely estimated zero-impact programs from our analyses.

One way to examine the sensitivity of cost-effectiveness estimates to significance level is as follows. Rank the programs based on their relative cost-effectiveness using the point estimate

of their impact and then re-compute the cost-effectiveness using the lower and upper bounds of the impact estimate. If this results in a program's cost-effectiveness changing drastically (for instance, moving from the top quartile to the bottom quartile of cost-effectiveness), then that cost-effectiveness estimate cannot be reported with as much confidence. This kind of check may be difficult to perform, however, in cases where only a few programs have sufficient data to be included in an analysis, making the range of cost-effectiveness estimates quite small.

Depending on the audience, it may be possible to include a confidence interval for cost-effectiveness based on some chosen level of significance, giving a sense of the precision of the estimation as well as the size of the impact. However, including a discussion of statistical power in the primary results to a non-research audience can obscure the main message of the analysis and make it harder for that audience to understand. It can also provide a false sense of precision. The true error bands around a cost-effectiveness calculation stem not just from imprecision on the estimated impact of the program, but also on estimates of costs and how they could vary between contexts. In some cases the uncertainty in cost-effectiveness analysis may stem more from the cost side than from the impact side, and it may be more appropriate to use error whiskers on the cost-effectiveness bar graph to highlight this variability than to focus on impact uncertainty. However, attempting to include error bands both for impact and costs is likely to be too confusing to be useful for many policymakers. It is straightforward to include information about the precision of estimates of impact at a secondary level for the more sophisticated reader, an approach J-PAL is starting to adopt on our website, where we will report confidence intervals of cost-effectiveness estimates based on the 90 percent confidence intervals around the impact estimate for each evaluation included. An example is included in Table 2.

CONFIDENCE INTERVALS OF COST-EFFECTIVENESS: ADDITIONAL YEARS OF EDUCATION PER $100 SPENT
RANGES BASED ON 90% CONFIDENCE INTERVAL OF PROGRAM IMPACT

Abdul Latif Jameel Poverty Action Lab (J-PAL)
www.povertyactionlab.org

	PROGRAM	COUNTRY	TIME FRAME	LOWER BOUND	PT. ESTIMATE	UPPER BOUND
1	Information Session on Returns to Education, for Parents	Madagascar	1 year	1.1	20.7	40.3
2	Deworming Through Primary Schools	Kenya	1 year	5.7	13.9	22.1
3	Free Primary School Uniforms	Kenya	1 year	0.33	0.71	1.10
4	Merit Scholarships for Girls	Kenya	3 years	0.02	0.27	0.52
5	Girls' CCT (Min. Amount)	Malawi	2 years	0.03	0.09	0.16
6	Girls' CCT (Avg. Amount)	Malawi	2 years	0.03	0.07	0.12
7	Girls' UCT (Avg. Amount)	Malawi	2 years	0.00	0.02	0.04
8	Iron Fortification and Deworming in Preschools	India	1 year	0.10	2.7	5.3
9	Camera Monitoring of Teachers' Attendance	India	–	NO SIGNIFICANT IMPACT		
10	Computer-Assisted Learning Curriculum	India	–	NO SIGNIFICANT IMPACT		
11	Remedial Tutoring by Community Volunteers	India	–	NO SIGNIFICANT IMPACT		
12	Menstrual Cups for Teenage Girls	Nepal	–	NO SIGNIFICANT IMPACT		
13	Information Session on Returns to Education, for Boys	Dominican Republic	4 years	1.0	3.1	5.2
14	PROGRESA CCT for Primary School Attendance	Mexico	4 years	0.02	0.03	0.04

2.4. Spillovers. In many cases, the effects of a program may spill over onto the untreated population, as in the provision of deworming drugs to schoolchildren to promote school attendance in Kenya (Miguel and Kremer 2004). Intestinal worms are spread through skin or oral contact with soil or water contaminated with infected fecal matter, and reducing the overall number of community members infected with worms has positive externalities in reducing local disease transmission to untreated children. 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 the transmission of worm infections. In the case of school-based deworming in Kenya, there were two kinds of spillover effects: the worm burden was reduced among children in treatment schools who chose not to take deworming drugs and also among children in nearby control schools which did not receive drugs that year.

In deciding whether to include the effect of spillovers in a cost-effectiveness analysis, one must assess whether spillovers would take place even when a program is scaled up. In the case of deworming, it is reasonable to think that not all children would be at school on the day of a large-scale deworming campaign or some others may not agree to take the pill, yet both groups would benefit from the within-school spillovers that would still occur due to the lower overall infection rate. Therefore, if a scale-up is likely to have imperfect coverage within the target population, then it is reasonable to include spillover effects that accrued to untreated targeted children in the original evaluation. However, a scale-up would probably attempt to reach all schools in an area, so spillovers to control schools would not be included in the calculation of benefits. In short, spillover effects should be included only when they are carefully measured and would also occur when the program is scaled up.

2.5. Aggregating Impacts. In its simplest form, calculating the total impact of a program follows this formula:

Total Impact of Program = Impact (per unit) × Sample Size × Program Duration

This calculation produces a figure for the total impact that a program had (time discounting is discussed in a later section). When there were differential impacts on different proportions of the population and those impacts occurred at different times, then this calculation requires more work. Any impact number used in calculations must correspond to the sample by which it is

being multiplied, and particular care must be taken when separating Treatment on the Treated (ToT) and Intention to Treat (ITT) effects. So long as the effect is multiplied by the correct sample, then ITT and ToT coefficients should give the same estimates of aggregate impacts, but it is important that no matter which estimate is used, the costs must always be aggregated over the entire population that was targeted.[5]

Another issue is that of proximal 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 is for a given expenditure. We therefore make a distinction between the proximal "success" of a program (immediate outcomes) and its final "impact" (effects on problems such as low learning, disease, etc.), resulting from immediate outcomes, which is a result of that proximal success. While most studies report final impact numbers, some only report proximal impacts. For example, studies

[5] Often cost-effectiveness is obtained by calculating the cost per beneficiary and then dividing this by the impact per beneficiary to get a cost-effectiveness ratio per beneficiary, without aggregating total impacts or costs. While mechanically this should give the same result as first aggregating costs and impacts across all beneficiaries and then dividing out those totals, we have chosen to begin with aggregate estimates, as it allows us to spell out the assumptions explicitly. For instance, if a remedial education program that cost $15 per child per year increased test scores by 0.15 standard deviations per child, it would appear relatively simple to divide these out. But if the impact per child had been measured after two years of the program, while the costs were incurred in both years, then the timeframes would not match. And if the cost per child had been simply summed up then divided by the number of children, then the costs from the second year of implementation would have been implicitly assigned a 0% discount rate. While aggregating impacts and costs and then dividing them out cannot prevent errors or accidental assumptions of this kind, the process of aggregation makes these issues more visible and provides a convenient opportunity to address them.

that promote chlorination of water to prevent diarrheal disease may measure success in terms of the number of additional households with chlorinated water (proximal success), but the ultimate objective of the program is to reduce diarrheal disease (final impact). If there is relevant evidence from other rigorous studies, especially meta-studies, to link the proximal impacts to final impacts, then this can help translate proximal effects into calculations of their final impact.

3 Quantifying Costs

Quantifying the costs of a program can appear deceptively simple, particularly when only aggregate cost data (such as the entire budgetary total) are reported in academic papers without a full explanation of what that budget includes and over what time period. But in order to ensure comparability across studies, it is necessary to obtain far more detailed cost data, to better understand the actual structure of the program and how its costs were distributed across beneficiaries and over time. To calculate the costs of many programs on a comparable basis, a number of judgments need to be made about what constitutes a "cost to the program." What should be included will depend on what the cost-effectiveness analysis will be used for, and by whom, but we will focus on a general principle for cost analysis given the perspective we have described—that of a policymaker allocating resources between different programs. In general we have taken the position that it is most useful to assess the marginal cost of adding a new education program, assuming that many of the fixed costs of running a school system will be incurred with or without the program (i.e. there is a "comparator case" against which the program is being compared).

3.1. Gathering Cost Data. Very few evaluations report comprehensive cost estimates, so cost data can be surprisingly difficult to obtain. Budget information tends to be incomplete and,

in some cases, inaccurate (Levin and McEwan 2001). The Ingredients Method is a useful way of making sure that all the appropriate costs have been included. Specifying all the ingredients necessary to replicate the program and then gathering unit cost information helps to ensure that a complete picture of the program's costs is included and guarantees comparability between programs. The academic papers on the relevant evaluation are usually a good starting point for the specification of ingredients, since they tend to provide an extensive description of the program itself. However, it is nearly always necessary to go back to the authors of the original evaluation and field staff to get clarifications on costs, how they were calculated, and how they broke down into different categories, as well as to get data on additional costs that were not listed in the academic paper. We are currently developing a general worksheet for researchers to use as a template for collecting cost data as they run evaluations. This will not only make it much easier to perform future cost-effectiveness analyses, but will also help improve comparisons of the different programs as part of cost-effectiveness analysis. The current iteration of J-PAL's worksheet is available in Appendix C, and we hope to refine it based on feedback from researchers and practitioners.

3.2. Modeling Program Costs at the Margin. When adding up the gathered ingredients costs, one must have a clear concept of what is meant by "the program" and the context in which it is assumed that this program will be replicated. Many evaluations examine different variations of an existing program, or evaluate a completely new one, so it is important to be cognizant of the starting situation against which the new cost model is being compared (the "comparator case"). Take, for example, an analysis of a computer assisted learning program in a school that already had computer facilities (Banerjee, Cole, Duflo, and Linden 2007). In the case of this evaluation, it was not necessary to pay for computers since they were already present in the school. However, if the program were to be scaled up to schools without computers, the cost of

the program would have to include the cost of setting up a computer lab. Alternatively, a school could have a lab that was unable to accommodate additional users, necessitating the purchase of more computers. In essence, these issues boil down to estimating the marginal costs of lumpy inputs such as hours of teacher time or computer use in the presence of discontinuous marginal cost functions.

This situation exemplifies how the right approach depends on the precise question being asked. The head of a school district that had underused computers would want to know the cost-effectiveness without including the cost of computers, while the head of a school district without any computers at all would want to have their costs included in the estimate. Without knowing precisely who is going to use our analysis we have to make an assumption about who is most likely to use the information. In this case, most schools in poorer regions of developing countries still do not have computers, and because the policy question that is often asked is whether or not computers should be provided, we include the costs of computers in our analysis.

On the other hand, take the example of a program providing merit scholarships to public school students based on standardized tests (Kremer, Miguel and Thornton 2009). In this case, the verification of test scores and selection of winners may be undertaken by school administrators whose salaries would have been paid even if the program had not taken place, and because the additional work is not very time consuming it seems likely that most government schools in which this program would be replicated would have the administrative capacity to select the top 15 percent of tests. It therefore seems reasonable to ignore the cost of administrators (but not of computers) because the analysis assumes that administrators (but not a computer lab) would already be present in most contexts in which a similar program would be replicated.

This point about what is reasonable to assume in a replication context is, in essence, another way of specifying the assumed situation into which a marginal program is being introduced (the comparator case). Cost-effectiveness analyses are not comparing the implementation of merit scholarships to doing nothing at all; if that were the case then every single cost associated with running the school in which scholarships were provided could be attributed to this program. The cost-effectiveness of a program is calculated as the marginal change in test scores (for example) as a result of the program, divided by the marginal change in costs because the program was implemented.

$$\textbf{Benefit Cost Ratio} = \frac{\textbf{[Test Scores with Program]} - \textbf{[Test Scores without Program]}}{\textbf{[Costs with Program]} - \textbf{[Costs without Program]}}$$

Underlying all cost-effectiveness calculations is an implied basic level of costs (teacher salaries, classroom supplies, etc.) and achievement (student attendance, test scores, etc.) that would exist even in the absence of the program. We call this the "comparator case" for cost-effectiveness analysis. Within this framework, the choice to "ignore" the cost of teachers who would still be paid in the absence of the new program is another way of saying that this cost appears in both terms of the denominator of the benefit-cost ratio above ("cost with program" and the comparator case of "costs without program"), and thus cancels out of the calculation. The context in which one assumes that replications would be implemented is merely another way of expressing the comparator program against which the new program is being evaluated, or the existing situation onto which this marginal program is being added.

Because of the important role that this comparator case plays in cost-effectiveness calculations, it is extremely important that it be well-specified. We cannot assume different

comparator cases for different programs in the same cost-effectiveness analysis. An example of this is the evaluation of a contract teacher program in Kenya (Duflo, Dupas, and Kremer 2010). This program gave funds to schools to hire a contract teacher, allowing them to split their first-grade classes into two, one taught by the contract teacher and the other by the regular civil service teacher. In practice, this program involved both decreasing class size and introducing a contract teacher who was accountable to the local parents, but the evaluation also allowed for comparison between contract and civil service teachers independent of class size (since both taught smaller classes).

Accordingly, one could estimate the costs of this program as the cost of hiring a contract teacher to allow for two first-grade classes, or as the costs of replacing civil service teachers with contract teachers. This choice has a significant impact on the outcome of cost-effectiveness analysis. The additional cost of adding one new contract teacher to a school while keeping everything else the same would include the new teacher's wages and training as well as any materials necessary to supply and oversee an additional classroom. But if a contract teacher was hired in place of a civil service teacher, the marginal cost of this new program would include the wages of the new contract teacher, net of the saved money that no longer needs to be paid to the old one. And because contract teachers have far lower wages than their civil service counterparts, the cost of replacing one government teacher with a contract teacher would actually be negative, resulting in negative cost-effectiveness (money saved per standard deviation increase in test scores). It is plausible to calculate cost-effectiveness for either (or both) of these programs, but it is important to be explicit about the structure of the program and the situation to which it is being compared.

Many development programs also require certain survey data during their implementation, but it

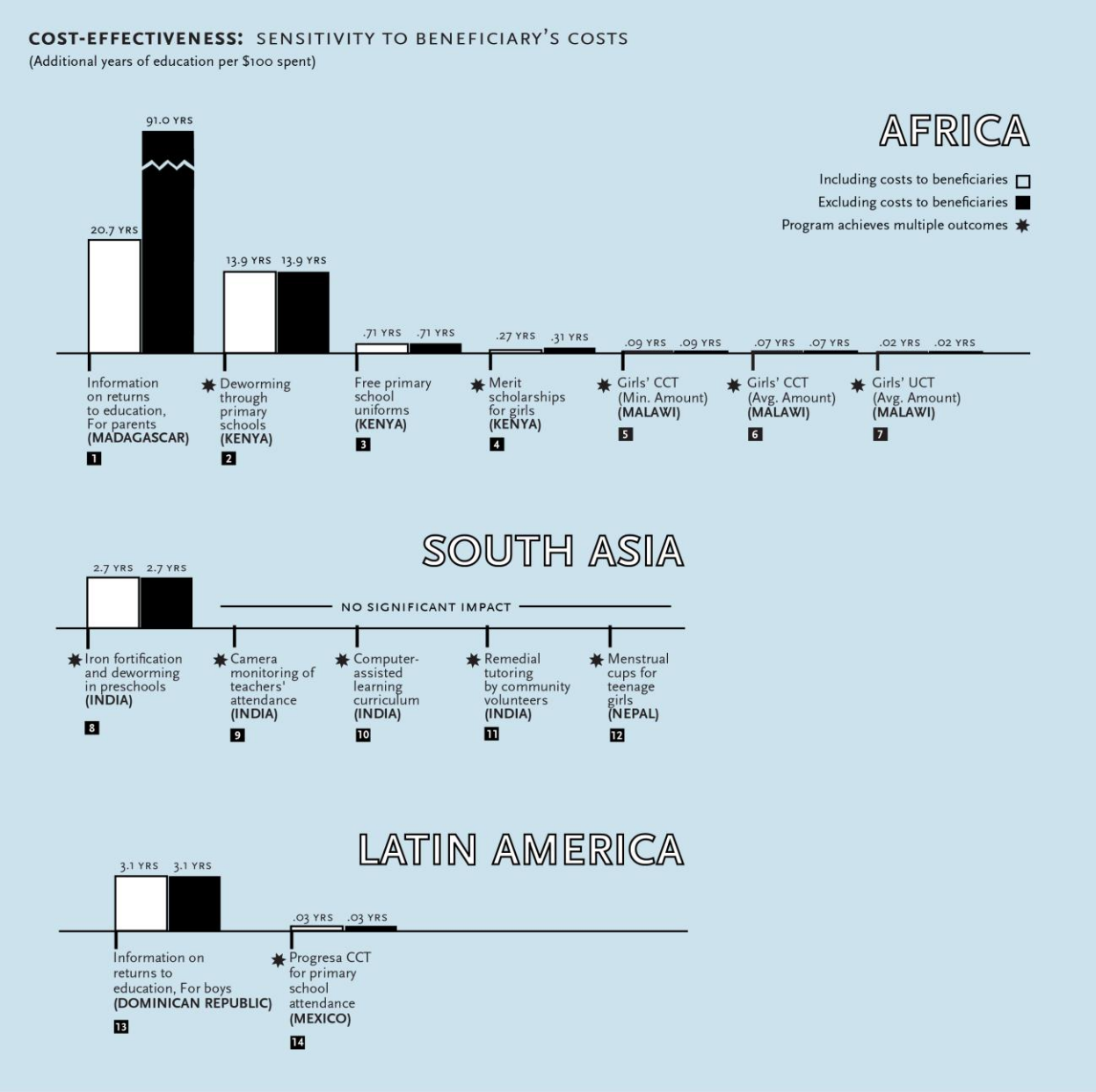
is not necessarily clear similar survey data would be available if the program were to be scaled up or replicated in a different context. For example, a program giving families information about returns to education relies on local wages for various education levels, and a conditional cash transfer program which uses detailed data from proxy means testing to identify beneficiaries. Even if an evaluated program was able to take advantage of existing information about the returns to education in a particular context, or a survey of the poorest households in a community, and thereby not incur additional survey costs it is still necessary to consider whether such information would exist in a typical context where such a program could be replicated. It is possible to do cost-effectiveness calculations either way—with or without the costs of collecting such underlying information that is critical for program implementation—but it is important to be explicit about what key program components are, and be consistent about the survey data that is assumed to be available in the context where the program may be replicated.

3.3. Goods and Services Procured at No Cost. In some evaluations, certain goods and services are provided at no cost to the program implementers—for example, a community may donate their labor to a project, or an outside organization may donate an input such as textbooks or deworming medicines. If the object of a cost-effectiveness analysis is to look at costs to society as a whole, the market cost of such free goods and services should be included. And even from the perspective of a particular implementing agency, inputs that were available for free on a smaller-scale project may not be available at no cost if the program is scaled up elsewhere, suggesting that the market costs of these free goods and services should be included. This process is relatively straightforward for material inputs that are necessary for the intervention (such as donated textbooks) for which the standard ingredients method can be applied using a market cost for the ingredient. In cases where services, such as labor for water source improvement projects, were provided at no cost by beneficiaries, the cost can be estimated as

what it would have cost to get the same work done by a paid laborer.

3.4. Costs to Beneficiaries. In many cases, programs also require beneficiaries to spend time contributing to the program—for instance, when parents must attend meetings to get information about the returns to education or give consent for the administration of deworming drugs. Some donors and policymakers may not be concerned with the costs of this time because it does not constitute an accounting cost to them. But because this time is a requirement of the program and represents a real cost to the user, we have chosen to include such costs wherever programs required users to commit their time. Where user costs are not a direct requirement of the program (for instance, in the girls' merit scholarship program in which parents were invited but not required to attend an awards ceremony) we do not include them as costs to the program.

Most evaluations report the average household income of the treatment and comparison group and we use this data to estimate the cost of users' time spent on the project. Because the average local wage rate for the poor in developing countries is quite low compared to total program costs, the relative ranking of the various programs in our cost-effectiveness analysis for student attendance does not change under differing assumed costs of foregone labor, as can be seen from the two charts below. However, the magnitude of estimated cost-effectiveness of some programs does change, especially those such as the information campaign in Madagascar, that had relatively low costs of implementation before including users' costs (see Figure 2).



3.5. Ingredients with Overlapping Uses. Many educational interventions require inputs such as teacher time, use of facilities, or administrative overhead. These ingredients are clearly necessary components of the program (it would be hard to adopt a new curriculum without a teacher to teach it), but there are sometimes reasons not to include every possible ingredient cost into an estimation of the cost-effectiveness of a program. In the earlier example of administering girls' merit scholarships in Kenya, we concluded that it was not reasonable to include the cost of

administrator's time because it overlapped so heavily with the basic functioning of the school.

On the other hand, it is also possible that other, future programs could piggyback onto the program one is examining now. For example, a significant proportion of the costs of PROGRESA come from the "targeting activities," where the poorest areas and households are identified for inclusion into the program. The information gained from these targeting surveys can be used by other programs in the future to identify beneficiaries, or even as simple demographic information to guide policymaking. Because there is no way to identify which of these costs may be distributed among other programs in the future, and because they still represent an accounting cost to the organization implementing the program at the time when the original program is run, we have not attempted to exclude such costs that may overlap with other programs in the future. Further, if a program such as PROGRESA is replicated in other countries where a targeting survey has already been conducted for other reasons, and/or the intended beneficiaries have already been identified, such survey costs would not be incurred. But again, since it is impossible for us to know this a priori, we have included these costs. When reviewing whether to replicate a program, implementers will be able to easily redo the cost-effectiveness analysis by excluding these costs if such a survey already exists.

A special case of the problem of "overlapping uses" is about inputs that are not completely used up in a year, such as school buildings or teaching materials which can be used for a longer period of time than is modeled in the cost-effectiveness analysis. If a program's impact is only measured over one or two years, but costs are included for goods that can actually continue to be used over many more years, it can result in an underestimation of the cost-effectiveness of the program. Consider a program which incurs large initial costs of procuring textbooks. The evaluation runs for one year, while the textbooks last for three years. If the full

cost of the textbooks is attributed to only one year of impact then the program's cost-effectiveness could suffer relative to another program that requires lower initial investment, but has higher variable costs. One way to deal with this problem is simply to use the rental cost for any goods that could be rented rather than purchased. However, it may not be possible to find rental costs for particular goods, such as investments in improving a building, in which case the cost of the input can be amortized over the assumed life of that good.

3.6. Transfers. 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. If we are concerned with costs and benefits to society as a whole, we should not include transfers as a cost. However, another way to look at the issue is that transfers are an example of a multiple-outcome program where one of the benefits is increased cash for the poor. If cost-effectiveness analysis focuses on one outcome only and ignores all other outcomes, why should cash outcomes be treated differently from, say, nutritional benefits? We will examine this question in some depth.

Mexico's PROGRESA program, where the government transferred money to families conditional on their children's attendance at school and healthcare check-ups, is a well-known example of cash transfers. The government's costs in this case can be divided into administrative costs (e.g. the costs of targeting poor households, 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). Administrative costs are a resource cost—real resources are used up by the program. Transfer costs are not a resource cost: the total resources in the economy do not go down, they simply get redistributed from one person to another. If we want to include PROGRESA in a cost-

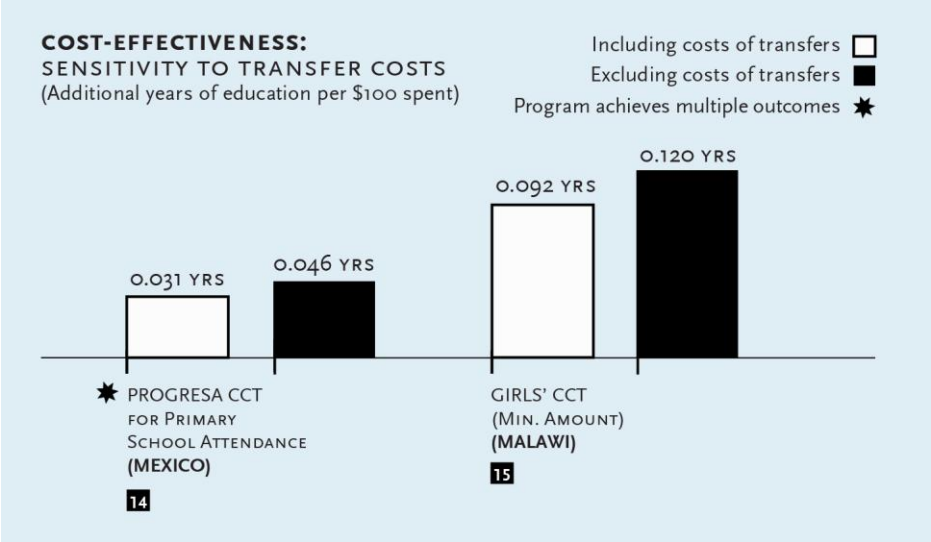
effectiveness analysis of alternative approaches to increasing attendance at school, should we include transfer costs or not?

If we were doing a cost-benefit analysis this would not be an issue. We would include all the costs to the implementer, including the transfers, as a cost but the cash received by the family would be included as a benefit, and the two would cancel each other out. (To be fully accurate we would want to include the deadweight cost of raising taxes to fund the subsidy, but we will ignore this as there are few good estimates of the deadweight cost of taxation in developing countries, and these are likely to vary considerably between countries with different tax systems.)

Within the context of a cost-effectiveness analysis, however, there are two ways to see this question, and they point to different answers. If we are interested in assessing cost-effectiveness to society as a whole, then transfers should not be considered a cost as they are not a cost to the society (except the deadweight cost of taxation). It could be argued, however, that a transfer is a cost to the implementer and a benefit to the beneficiary. Conditional cash transfers achieve more than one outcome—for example, they increase school attendance and they redistribute cash to the poor. Through this lens, it becomes difficult to see why we should adjust for the benefits of cash transfers when we don't adjust for other outcomes—from the nutritional benefits of school meals, for example, or to the health benefits of deworming. We argued that we did not want to put a monetary value on these benefits as they will vary across contexts, and cost-effectiveness estimates are likely to be sensitive to the choice of values. What is the monetary value of a child receiving a free meal at school? Is it the full value of the meal? It is probable that the child's family would not value it at the full cost of the meal. If given the cash equivalent of the meal they would probably not spend it all on food for the child. One reason to

give in-kind benefits is because one thinks that the implementer has a different valuation of benefits than the recipient and one wants to skew their spending in a particular direction, in this case towards child nutrition. That said, the family would undoubtedly put a positive value on the meal (Kremer and Vermeersch, 2004, actually attempt to back out families' implicit valuation of the meals). Indeed, the majority of programs included in the cost-effectiveness analysis of school enrollment and attendance include multiple benefits, and some of them are monetary. Assessing the appropriate valuation of all these benefits, however, is highly problematic, and in our view makes the analysis extremely opaque. An assessment of the cost-effectiveness of the school meal program, for example, would be very sensitive to exactly how much one thought people valued the meal.

Our conclusion is that a cash benefit is another case of a multi-outcome program, but it is a special one because it is easier to estimate its value. However, even in the case of cash transfers it is not necessarily true that the marginal benefit of $1 to a poor household is equal to the marginal value of $1 to a wealthier household. Therefore, we attempt to clearly show what proportion of costs is due to transfers and sometimes show cost-effectiveness without transfers. Figure 3 illustrates the impact of excluding cash transfers on the cost-effectiveness of PROGRESA. To compute the cost of transfers, we used the disaggregated transfer amounts that are linked only to primary school attendance and not transfers associated with either secondary school or health outcomes.



A practical point to keep in mind when deciding whether to include transfers as costs is that while funders do care about costs incurred and benefits received by beneficiaries, they also have budget constraints that require accounting for both administrative costs and transfer costs. They may want to know how much "bang they can get for their buck" in terms of impact on a narrow outcome, such as school attendance, from different programs—including conditional cash transfers, and netting out transfers makes it hard for them to do this calculation. We therefore provide a version of the cost-effectiveness analysis with transfers included as a cost because funders do face budget constraints and because other programs also have multiple outcomes.

3.7. High-Level Management Overhead. One of the most difficult cost items to incorporate in cost-effectiveness analysis is the incremental cost of indirect overhead. This is because the additional time, effort, and cost of high-level administration and oversight that is incurred by the organization due to a new program is rarely tracked or quantified. This section is not meant to revisit the previous category of costs of a new program that overlap with the basic functioning of a school or local administration (such as paying for facilities or electricity). Rather, it will focus on the costs of higher-level overhead, such as additional administration time

needed to process the payroll of new employees, or the time an existing civil servant spends overseeing the implementation of the program. These additional costs are almost never reported, especially at the pilot or research stage when they represent a small amount in a (likely) much larger organizational budget. As such, they are almost impossible to observe and any estimations by us would be extremely imprecise. In many cases, the costs of such high-level overhead are likely to be relatively small compared to the other costs of the program, and are also likely to be similar across the various programs being compared. This suggests that, in most cases, they can be netted out of calculations without biasing the relative cost-effectiveness estimates.

If there were a reason to believe that programs within an analysis had drastically different costs of indirect overhead (for instance, because a program would require protracted renegotiation of union contracts by very senior management, such as some of the programs described in Chapter 6), then it could be possible to put together some estimate of the indirect overhead costs. Assuming that the indirect overhead costs are a function of the amount of personnel costs, one could assume an additional 10 percent or 15 percent of cost for the indirect overhead of administration. However, the choice of an overhead "rate" would be extremely arbitrary, and we do not make these assumptions in our analysis.

3.8. Experimental vs. Scalable Modes. 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 have the potential to increase the cost-effectiveness of programs. On the other hand, there may be disadvantages to working at scale, such as the increased difficulty of administering a program over a wide area or the cost of hiring new senior management to administer the scaled program, which may affect both costs and impacts. The ratio of fixed to variable costs can also

impact how cost-effective a program looks at pilot versus at scale. When this ratio is high, a program will not look as cost-effective because the fixed cost is spread over only a small number of beneficiaries. But a program with high fixed costs relative to its variable costs may be more cost-effective at scale, when the fixed cost is spread over a larger number of beneficiaries. When scale economies are very obvious and guaranteed to be realized, these may be used to estimate the cost of a program.

For example, in performing a cost-effectiveness analysis of a program that supplies flipcharts to schools in Kenya, the budget would report the cost of flipcharts based on their purchase through retail outlets. If this program were to be adopted across an entire state, flipcharts could be purchased in bulk, and so bulk costs should be applied to this ingredient. This can be particularly important in programs where the majority of the costs come from goods or services that are particularly sensitive to scale. However, it is important to be cautious in the application of "scale economies" to the ingredients in the program. If a program has not been tested at scale there may be a concern that while the costs would go down with scale, so might the quality of monitoring or delivery, and hence the impact. Without good reason, it is often better to stick with the actual costs and actual benefits of the program as it was evaluated.

For this reason, wherever changes have been made to an original program design in its 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 is advisable to perform a detailed survey of local costs before choosing to expand the program. This is what J-PAL advises policymakers who are looking to expand or replicate programs found to be successful in evaluations. As more and more piloted evaluations are scaled up in the next few years, there will be a better understanding

among researchers and policymakers about how individual costs of goods and services in evaluations translate into costs in large scale-ups.

4 Using Common Units

One of the unseen challenges of cost-effectiveness analysis, beyond the selection of an appropriate outcome measure and the inclusion of the appropriate costs, is converting all costs and impacts into "common units" adjusting consistently for inflation, exchange rates, and year of implementation. For the sake of clarity, it is useful to define two terms at the beginning of any analysis:

> Year of Analysis: The year in which a cost-effectiveness analysis is undertaken or the choice between the various programs is made. The year of analysis must be consistent for all programs in an analysis. If the year of analysis is 2010, all final cost figures should be inflated to 2010 dollars.
>
> Base Year: The year in which the program being evaluated was launched. Before inflating forward to the year of analysis, costs and benefits are discounted back to the base year of the program so inflation is compounded over the correct number of years.

It is necessary to define a single year of analysis that is used for all programs in a given analysis. If this is not done, and the base years of each individual program are used as the effective year of analysis (i.e. costs are reported in terms of that year's currency), then differences in costs may be driven by inflation between the different years of analysis. When adjusting for common units, costs are first converted into a common currency (usually the United States dollar), then converted to prices in terms of Base Year dollars, and thereafter the present value of these cost

flows in the Base Year is computed. Costs are then inflated forward to their value in the Year of Analysis using a common inflation rate.

4.1. Adjusting for the Base Year of the Program. When a program's costs and impacts are distributed across time, it is necessary to discount them back to their present value in the base year of the program to account for an organization's time preference for both costs and benefits. There is no universally applicable real discount rate in the literature, and in practice there are significant variations in public discount rates applied by different countries. Developing countries tend to apply higher social discount rates (8–15 percent) than developed countries (3-7 percent) (Zhuang et al 2007). The "correct" discount rate depends on who is making the investment: different decision makers will use different methods to estimate their discount rate.

The discounting of costs is representative of the choice a funder faces between incurring costs this year, or deferring expenditures to invest for a year and then incurring costs the next year. An organization or government's discount rate is usually calculated as the social opportunity cost of capital (SOC). This rate varies across countries and organizations, but there seems to be a higher variance in the public, rather than private, cost of capital.

The discounting of benefits, on the other hand, represents how an end user of the program would trade off between the uses of the services this year versus next year. The appropriate discount rate for such a calculation is the social rate of time preference (SRTP), or the rate at which users would trade off one unit of consumption today versus one unit of consumption tomorrow. There is relatively little information on the time preferences of people in poorer countries, and the fact that variations will depend upon the intended user of the program, rather than the implementer, makes it difficult to choose one rate which would be applicable in a variety of cases.

If an organization were performing a cost-effectiveness analysis of programs that they run in particular countries, then it would be possible to use the SOC to discount their costs knowing their own cost of capital and use the SRTP of the country in which beneficiaries live to discount effects. However, in performing general cost-effectiveness analysis that is likely to be used by policymakers in different organizations and countries, one is unlikely to have such specific information about users, and so it is practical to choose a single discount rate. Because of the high variance and scarce empirical data on time preferences in the developing world, the SRTP is not a practical option. This suggests that the SOC may be the best available discount rate, but the question remains as to which country or organization's SOC should be used. International aid tends to come from the developed world (even when it is channeled through local governments), and so the opportunity cost of devoting capital to a given program is most often based on the foregone return or cost of borrowing on the developed country capital markets. For a list of discount rates used by various governments and organizations, see Table 2. One of the most striking features of this table is the relative similarity of rates across organizations using the SOC to calculate their discount rate. Looking at the median rate of countries using the SOC methodology suggests that 10 percent is a reasonable rate for discounting the costs and benefits of educational programs in developing countries.

Table 2: Survey of Social Discount Rates[6]

Country	*Discount Rate*	*Theoretical Basis*
Germany	*3%*	*Based on federal refinancing rate*
Norway	*3.5%*	*Unknown*
United Kingdom	*3.5%*	*SOC until early 80s, SRTP after*
France	*4%*	*SRTP approach*
Spain	*4-6%*	*SRTP approach*
Italy	*5%*	*SRTP approach*
United States (OMB)	*7%*	*Unknown*
People's Republic of China	*8% for short- and medium- term projects*	*Weighted average approach*
Canada	*10%*	*SOC approach*
New Zealand (Treasury)	*10%*	*SOC approach*
Asian Development Bank	*10-12%*	*Unknown*
India	*12%*	*SOC approach*
Pakistan	*12%*	*SOC approach*
Philippines	*15%*	*SOC approach*

[6] Source: Juzhong Zhuang et al "Theory and Practice in the Choice of a Social Discount Rate for Cost-Benefit Analysis". ERD Working Paper No. 94. Asian Development Bank, 2007.

Many of the programs included in our examples of analysis were run and evaluated over a relatively short time frame. Except in cases where there are large one-time start-up costs, most of these programs can be examined over a one or two year time frame, and so their comparative cost-effectiveness is not particularly sensitive to the choice of a discount rate. The relative insensitivity of this kind of estimate to changes in the discount rate is shown in Table 3.

SENSITIVITY OF COST-EFFECTIVENESS TO DISCOUNT RATE
ADDITIONAL YEARS OF SCHOOLING PER $100 SPENT

Abdul Latif Jameel Poverty Action Lab (J-PAL)
www.povertyactionlab.org

	PROGRAM	COUNTRY	TIME FRAME	5% DISCOUNT	10% DISCOUNT	15% DISCOUNT
1	Information Session on Returns to Education, for Parents	Madagascar	1 year	20.7	20.7	20.7
2	Deworming Through Primary Schools	Kenya	1 year	13.9	13.9	13.9
3	Free Primary School Uniforms	Kenya	1 year	0.71	0.71	0.71
4	Merit Scholarships for Girls	Kenya	3 years	0.260	0.271	0.273
5	Girls' CCT (Min. Amount)	Malawi	2 years	0.09	0.09	0.09
6	Girls' CCT (Avg. Amount)	Malawi	2 years	0.07	0.07	0.07
7	Girls' UCT (Avg. Amount)	Malawi	2 years	0.02	0.02	0.02
8	Iron Fortification and Deworming in Preschools	India	1 year	2.7	2.7	2.7
9	Camera Monitoring of Teachers' Attendance	India	–	NO SIGNIFICANT IMPACT		
10	Computer-Assisted Learning Curriculum	India	–	NO SIGNIFICANT IMPACT		
11	Remedial Tutoring by Community Volunteers	India	–	NO SIGNIFICANT IMPACT		
12	Menstrual Cups for Teenage Girls	Nepal	–	NO SIGNIFICANT IMPACT		
13	Information Session on Returns to Education, for Boys	Dominican Republic	4 years	3.32	3.11	2.93
14	PROGRESA CCT for Primary School Attendance	Mexico	4 years	0.032	0.031	0.031

*The discount rate is applied at the end of years, so programs examined over a one year time frame are not discounted.

In cases where there are large one-time costs, such as targeting activities or construction of new buildings, but benefits that accrue over a longer time frame, the cost-effectiveness estimates would be more sensitive to the choice of discount rates.

4.2. Adjusting for Inflation. When performing cost-effectiveness analysis based on the results of an impact evaluation, ingredient costs are frequently taken from the costs incurred in the evaluation itself, which in many cases are reported in terms of their nominal amounts in the year in which they were incurred. Depending on how costs are reported, this can necessitate up to two adjustments for inflation. First, any costs which were reported in terms of the year in which they were incurred (e.g. 2004, 2005, and 2006 USD) must be deflated back to their real value in base year dollars (2004 USD), to account for the fact that inflated prices may make later costs appear larger even if they are identical in real terms. Secondly, once the present value of the cost stream has been calculated from the perspective of the Base Year, it is usually necessary to inflate this figure forward to reflect what it would cost in the Year of Analysis (in our student attendance example, this is 2010 USD). For both of these calculations it is preferable to use the average GDP deflators 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.

On average there should be no difference between converting to dollars and applying the US inflation rate versus applying the local inflation rate and then exchanging currencies, but in practice distorted exchange rates may not always capture inflation adequately. For this reason we have chosen to convert to dollars and then use the US inflation rate. We follow the same methodology consistently in all J-PAL cost-effectiveness analyses.

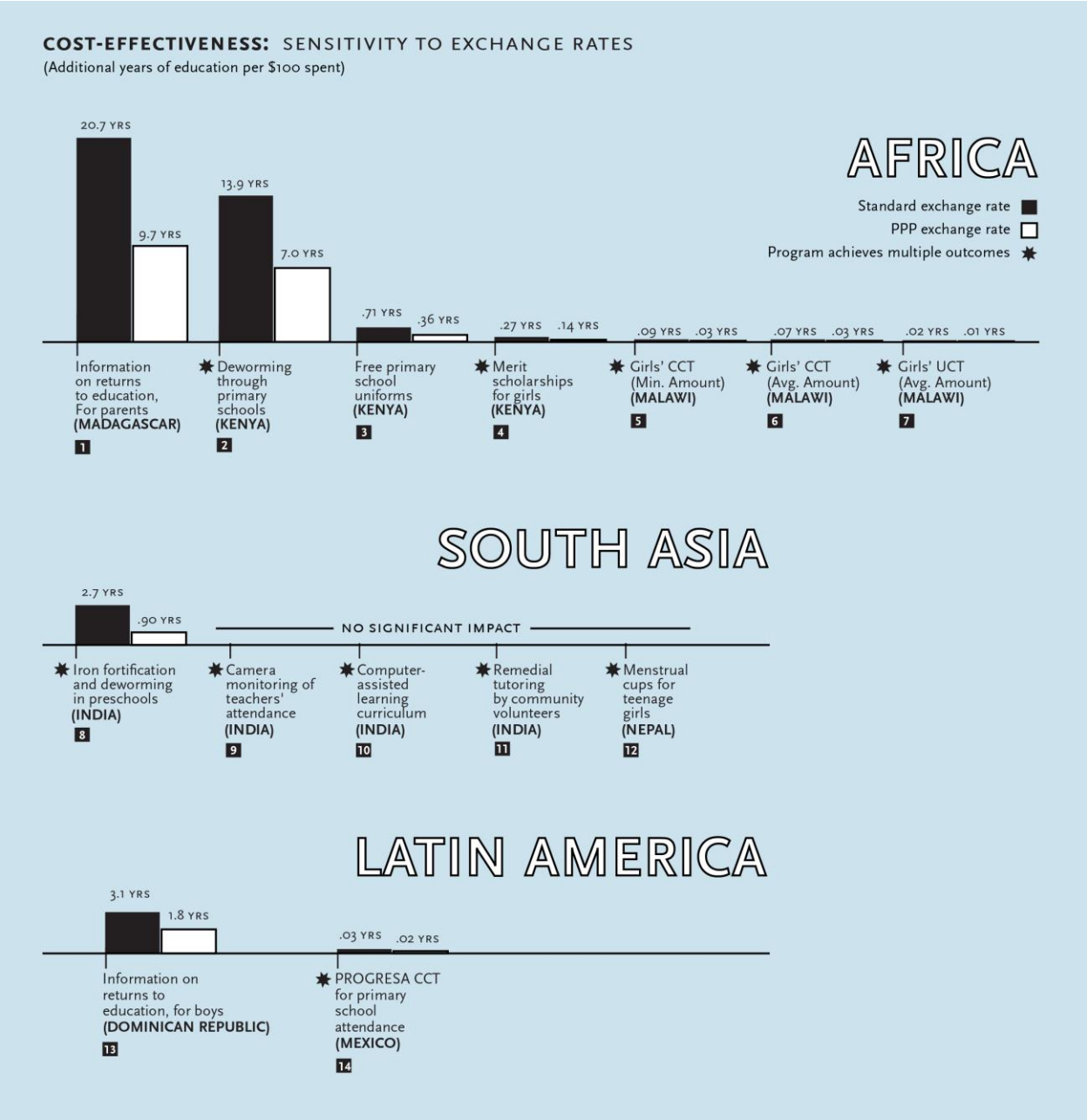
4.3. Currency and Exchange Rates. Many evaluations report program costs in US dollars (USD), but some also report costs in local currencies, and where costs are gathered from

a number of sources there may even be a mixture of units. It is obvious that all programs being evaluated must have their costs exchanged into a single currency, but the choice of an exchange rate has significant implications for the interpretation of the results.

When standard (i.e. market) exchange rates are used, the resulting estimates represent the cost-effectiveness of that program assuming the relative price levels for different goods in the country in which it was originally implemented. The difficulty in using standard exchange rates is that there are significant differences in the relative prices of different goods across countries. Purchasing Power Parity (PPP) exchange rates adjust somewhat for the different price *levels* in different countries, which are driven by the higher prices of non-tradables in wealthier countries. But since PPP is based on a standard basket of goods and services, it does not completely adjust for the different relative prices of the goods and services used in a particular program across countries, because of the variations in factor endowments across countries. For example, skilled labor is far cheaper in India than in Mexico, so if a program that is more intensive in skilled labor is piloted in Mexico, it will look less cost-effective than a similar program piloted in India. But there is a danger that some readers may not appreciate this fact and will assume that the relative prices of different categories of goods have been completely adjusted in the PPP version.

Moreover, because PPP exchange rates effectively adjust to what a program would cost in the United States ($1 PPP = $1 USD), the cost-effectiveness estimates for all of the programs will decrease significantly in absolute size, potentially giving an inaccurate estimate of what could be achieved with a given expenditure in a developing country. This could be resolved if readers were first converting from PPP back into standard US dollars using the conversion factor for each program's country of origin, but it is almost certain that this kind of mental calculation will be done straight from PPP to a policymaker's domestic currency using standard rates, while

U.S. based policymakers might miss the conversion altogether. Therefore the default presentation of our cost-effectiveness results uses standard exchange rates, but when possible, we will also present a version with PPP rates to show that the relative cost-effectiveness of the programs does not change (see, for example, Figure 4). Given that relative prices do differ across countries, it is useful before launching a large program to do an assessment of costs in the intended location, especially when costs in the original program were driven mostly by a particular factor cost, such as wages. Again, providing the underlying calculations should enable policymakers to make such adjustments.



4.4. Order of Operations. For simplicity's sake, we keep track of the units for costs and impacts, including currency, year, and whether present value has been applied. Table 4 specifies the order of operations that J-PAL uses to harmonize cost units for the most complex program in our example: a program for which there are cost data reported in the prices of the years in which the costs were incurred, in local currency. This particular order of operations is not necessarily better than any other, the important thing is that an order be selected and consistently applied to

all programs in an analysis.

Step	Operation	Unit of Currency (e.g.)
1.	Gather cost data.	2004, 2005, and 2006 pesos
2.	Exchange into United States dollars using the year-specific exchange rates.	2004, 2005, and 2006 USD
3.	Deflate nominal costs back to real value in Base Year (2004) prices, using average annual US inflation rate over time elapsed between Base Year and incurrence of costs.	2004 USD (incurred in 2004, 2005, and 2006)
4.	Take the present value of this cost stream using a 10% real discount rate.	PV of the cost stream in 2004, in 2004 USD
5.	Inflate forward to Year of Analysis (2010), using average annual US inflation rate over time elapsed between Base Year and Year of Analysis.	PV of the cost stream in 2010, in 2010 USD

5 General Issues with Cost-Effectiveness Analysis

5.1. 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 only attempt to measure 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 high returns at the outset, since they are the only ones with specialized knowledge. 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, as discussed earlier. On the other hand, benefits sometimes become larger as programs are scaled up and behavior change is reinforced by seeing peers undergo the same behavior change. The marginal benefits of education can also increase as more people become educated and there are complementarities between skilled workers. It is difficult 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. The level at which general equilibrium effects will be observed can vary in different situations and can only be determined empirically, although it may be possible to make reasonable estimates of how a program will perform at scale based on the design of the program and the size of the target population. It is not practical to attempt to include general equilibrium effects in our cost-effectiveness analyses, although we will attempt to flag the most problematic ones.

5.2. Initial Levels of Underlying Problems. Different countries or regions will have different intensities of the underlying problems that programs are seeking to address. For example, there is far higher baseline attendance in Mexico than in Kenya. Because of this, they may be at different points on the "marginal benefit curve" of intervention, which can result in variations in the cost-effectiveness figures for the same program piloted in different regions. For instance, a deworming pill should be equally effective at killing intestinal worms in Africa or in Asia, but there is declining marginal benefit to more deworming, and so the number of school days gained as a result of a deworming program depends on the pre-existing intensity of infection in any particular place. Thus deworming might buy five additional days of schooling in western Kenya, where worm prevalence is very high, but fewer additional days of schooling in Andhra Pradesh state in India, where the worm prevalence is lower. Similarly, intuition tells us that the "last mile problems" would make it harder to increase school enrollment from 90% to 95% than from 50% to 55%. But whether this is the case, and to what extent, is an empirical

question that is very hard to resolve a priori. In some cases, especially among education interventions, we have used impacts (such as an increase in test scores) reported in terms of standard deviations to mitigate some of the issues associated with the initial level of the underlying problem that the programs seek to address.

When considering the baseline rates of the initial problem that a given program tries to address, there are two separate issues: putting programs on more similar footing, and producing accurate estimates of what a program's impact would be where a user wanted to implement it. Our goal in comparative analysis is to generate good estimates of the relative cost-effectiveness of different programs, but these estimates will always be representative of programs as they were piloted in a particular context. Providing policymakers with the underlying calculations for cost-effectiveness analysis can allow them to run sensitivity analysis by adjusting factors that are important for their context, such as population density. However, given that most pilots are tested in areas where the underlying problem is severe (for instance, where parents systematically underestimate the returns to primary education) and programs are likely to be scaled up or replicated only in areas where that problem is also salient, it is unlikely that baseline levels in pilot areas will be significantly different from baseline levels in the replication context.

Another way to minimize the bias in comparisons across different contexts is to group programs by the region or type of country in which they were piloted. Programs in similar regions or national income brackets are more likely to have common elements in terms of the baseline rates of the underlying problems. Thus a policymaker may be able to study the programs that were piloted in their region, or in countries facing similar problems and that are at a similar stage of development to compare cost-effectiveness. For example, we have grouped our student attendance graph by the region in which the program took place.

While grouping programs by the region or income group of the country in which they were piloted reduces the bias in comparisons across regions, it can also make comparisons difficult when there are only a few programs in a particular region. Our student attendance and enrollment analysis is not yet complete, and more programs will be added to allow for better comparisons within regions.

5.1. Generalizability of Costs. 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 absence or disease can help to ensure the comparability of impact figures, there are other location-specific parameters that can influence the cost-effectiveness of a program. For example, in the area of India where the Balsakhi remedial tutoring program was tested, there were sufficient volunteers available with a high enough education level to take advantage of the pedagogical materials provided and use them to provide out-of-school tutoring for local children. In another state in India where education levels are lower, it might be harder to find a volunteer to teach in every village, and so the fixed cost of developing the pedagogical materials would be spread among a smaller pool of villages. These costs can vary across different contexts, even within the same country, due to demographic factors. For example, the number of schools an administrator can visit in a day may vary depending on school size, transportation infrastructure, etc.

Similarly, some other contextual factors, such as population density in the area in which a program is piloted, can influence estimates of cost-effectiveness. Many programs will cost more, and be less cost-effective, when the population is sparsely settled compared to more densely populated regions. But it should be relatively easy to adjust cost estimates based on expected population density if the amount of goods and services necessary per household or individual is

known. If all programs within a given analysis are piloted in areas with similar population density, then population density should not present any problems for their relative cost-effectiveness. If, however, different programs are piloted in areas with dramatically different population densities, policymakers may be interested in substituting the population density for the region of interest to them.

In addition to its complicating effect on comparisons, the sensitivity of cost-effectiveness estimates to certain parameters complicates the presentation of cost-effectiveness results. Simple bar graphs which can't include information on all sensitivities can 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 a cost-effectiveness analysis of the pilot program. While cost-effectiveness analyses are intended to provide a means of comparison between different programs if they were implemented by the policymaker, they are not intended to reflect exactly what a particular program would cost to implement in any setting. To reflect the way that costs per impact can be expected to vary as a function of certain parameters, it can be useful to select the most relevant variable for a given program and show how the cost-effectiveness of that program would change with that variable over a reasonable range. Using the example of a remedial education tutoring program, one can present the point estimate of how much it would cost to increase test scores by one standard deviation if the actual (observed) proportion of villages in the study (say, 75 percent) found a tutor and implemented the program, as well as a range around this point estimate of cost per unit increase in standard deviations of test score if 70 percent of villages implemented the program and if 80 percent of villages implemented it. For educational interventions, the largest cost item is often wages, which vary widely across contexts, and so it can be instructive to include an interval of cost-effectiveness under a reasonable range of wages.

As J-PAL performs cost-effectiveness analyses, more detailed spreadsheets will be made available to help those wishing to scale up and implement programs for which cost-effectiveness results are given in different contexts. Such dissemination of underlying calculations, thoroughly cited and explained, will allow other organizations to examine the underlying calculations and modify various parameters based on their situation. For example, a 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 to generally adapt the figures shown in original analyses to the needs of different users.

6 Conclusion

Cost-effectiveness comparisons can be a powerful tool to inform the debate about how best to improve education in developing countries. By placing program costs and impacts on a similar basis these analyses can make comparisons of different programs very salient. But a number of judgments need to be made in the process of undertaking cost-effectiveness analyses, for example, what is the appropriate discount rate or exchange rate to use, and should transfers be included as costs or not? In this chapter we have set out a particular set of assumptions or judgments which we believe provide a useful basis for comparing education programs in developing countries. As we have pointed out throughout the article, determining which assumption is best depends on the precise question or context to which the analysis will be applied. In some cases the cost-effectiveness calculations are not very sensitive to changes in the assumptions within reasonable ranges. This is the case for discount rates, for example. In other cases the absolute values change with different assumptions but the relative ranking of programs does not change (this is true for current exchange rates vs. purchasing power parity, for

example). Wherever the results are highly sensitive to a particular assumption we highlight this fact.

A cost-effectiveness analysis should be taken as one more input into a decision about which programs to fund, along with other considerations, and not the only factor. While costs and impacts may vary between settings, by making available the underlying data and calculations that go into the analysis, we hope to provide a framework for funders and implementers to think through what results they might expect in their particular context. We have found this useful for working with funding organizations to assess the likely cost-effectiveness of scale-ups in new contexts by adapting the inputs and methodology set out here.

None of this work is possible without detailed data on costs and impacts. The process of doing more and higher quality comparative cost-effectiveness work will be greatly enhanced if researchers record detailed cost information during their field research. We hope this chapter helps develop standard ways to collect data on costs and impacts. If a consensus emerges about the best way to measure education quality and quantity, it will make comparisons much more useful. In some subject areas, J-PAL is attempting to coordinate with researchers on agreed standardized outcome and cost measures, but in education as with other sectors, there is still more work to be done.

With all the assumptions and imperfections involved in undertaking comparative cost-effectiveness analyses, some may argue that they should not be undertaken. In our view, policymakers will always make comparisons across programs about cost-effectiveness - they have to, given limited resources and the large number of programs aimed at similar outcomes that compete for those resources. Providing policymakers with tools that are clear about the assumptions being made and that can be easily adapted makes the process more transparent, less

ad hoc and is likely to increase the use of rigorous research evidence in policymaking.

Bibliography

Anand, Sudhir, and Kara Hanson. "Disability-Adjusted Life Years: A Critical Review." *Journal of Health Economics*, 1997: 685-702.

Baird, S., Ozler, B. & McIntosh, C., 2011. Cash or Condition? Evidence from a Cash Transfer Experiment. *Quarterly Journal of Economics,* 126(1709-1753).

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122, no. 3 (August 2007).

Barnett, W. Steven "Benefit-Cost Analysis of the Perry Preschool Program and Its Policy Implications." *Educational Evaluation and Policy Analysis* 7, no. 4 (1985): 333-342.

Behrman, Jere R., and James C. Knowles. "Economic Returns to Investing in Youth," in Jere R. Behrman, Barney Cohen, Cynthia Lloyd and Nelly Stromquist, eds, *The Transition to Adulthood in Developing Countries: Selected Studies*, Washington, DC: National Academy of Science-National Research Council, 424-490.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/Oportunidades." *The Journal of Human Resources*, 2011: 93-122.

Bobonis, Gustavo, Edward Miguel, and Charu Puri-Sharma. "Anemia and School Participation." *Journal of Human Resources*, 2006: XLI:692-722.

Burde, Dana, and Leigh Linden. "The Effect of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan." 2010.

Caldés, Natàlia, David Coady, and John A. Maluccio. "The Costs of Poverty Alleviation Transfer Programs: A Comparative Analysis of Three Programs in Latin America." *FCND Discussion Paper No. 174*, February 2004.

Carnoy, Martin. "The Economic Costs and Returns to Educational Television." *Economic Development and Cultural Change*, January 1975: 207-248.

Coady, David P. "The Application of Social Cost-Benefit Analysis to the Evaluation of PROGRESA: Final Report." International Food Policy Research Institute, Washington, D.C., 2000.

Duflo, E., Hanna, R. & Ryan, S. P., Forthcoming. Incentives Work: Getting Teachers to Come to School. *American Economic Review.*

Duflo, Esther, Pascaline Dupas, and Michael Kremer. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." 2010.

Evans, David, Michael Kremer, and Mũthoni Ngatia. "The Impact of Distributing School Uniforms on Children's Education in Kenya." November 2009.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. "Teacher Incentives," American Economic Journal: Applied Economics, American Economic Association, vol. 2(3), pages 205-27, July,

2010.

Hummel-Rossi, Barbara, and Jane Ashdown. "The State of Cost-Benefit and Cost-Effectiveness Analyses in Education." *Review of Educational Research* 72, no. 1 (2002): 1-30.

Jensen, R., 2010. The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 25(2), pp. 515-548.

Kremer, Michael, and Christel Vermeersch. "School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation." November 2004.

Kremer, Michael, Rebecca Thornton, and Edward Miguel. "Incentives to Learn." *The Review of Economics and Statistics*, 2009: 437-456

Levin, Henry M. "Waiting for Godot: Cost-Effectiveness Analysis in Education." *New Directions for Evaluation* 90 (2001): 55-68.

Levin, Henry M., and Patrick J. McEwan. *Cost-Effectiveness Analysis,* 2nd Edition. Thousand Oaks, CA: Sage Publications Inc., 2001.

Levin, Henry M., Gene V. Glass, and Gail R. Meister. "Cost-Effectiveness of Computer-Assisted Instruction." *Evaluation Review* 11, no. 1 (February 1987): 50-72.

Linden, Leigh. "Complement or Substitute? The Effect of Technology on Student Achievement in India." 2008.

Lopez, Humberto. "The Social Discount Rate: Estimates for Nine Latin American Countries: Policy Research Working Paper 4638." Latin American and the Carribean Region, The World Bank, 2008.

Miguel, Edward and Michael Kremer. "Worms: Identifying Impacts on Education and Health in the Presence of Externalities." *Econometrica*, January 2004.

Nguyen, Trang. "Information, Role Models, and Perceived Returns to Education: Experimental Evidence from Madagascar." January 2008.

Nguyen, Trang, and Gerard Lassibille. "Improving Management in Education: Evidence from a Randomized Experiment in Madagascar." 2008.

Oster, E. & Thornton, R., 2011. Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics*, 3(1), pp. 91-100.

Schultz, Paul T. *Final Report: The Impact of PROGRESA on School Enrollments.* International Food Policy Research Institute, 2000.

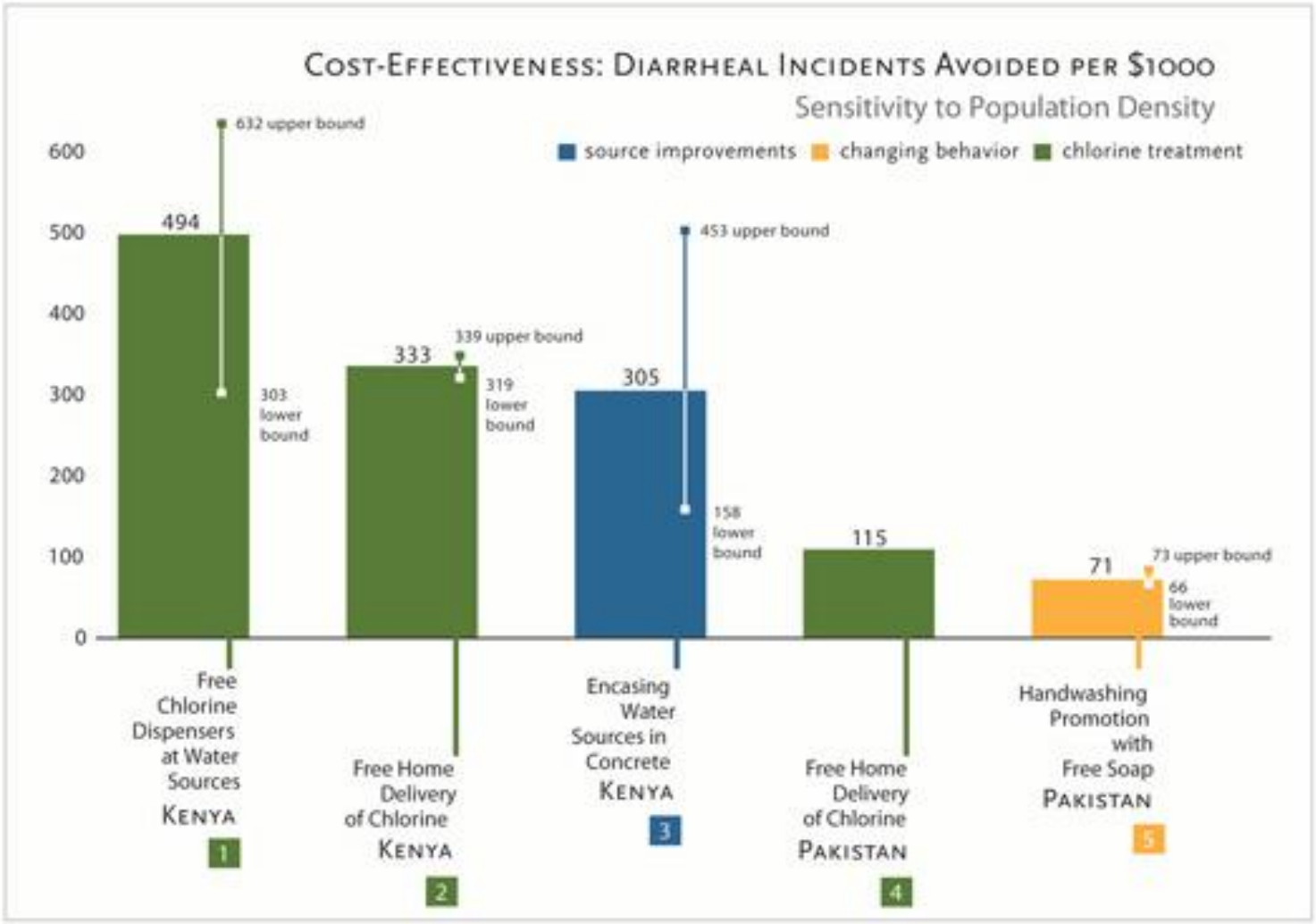
Schultz, T. Paul. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Program." *Journal of Development Economics* 74 (2004): 199-250.

Thornton, R., Kremer, M. & Miguel, E., 2009. Incentives to Learn. *The Review of Economics and Statistics*, 91(3), pp. 437-456.

U.S. Department of Health and Human Services. "Cost-Effectiveness in Health and Medicine: Report to the U.S. Public Health Service by the Panel on Cost-Effectiveness in Health and Medicine." 1996.

Zhuang, Juzhong, et. al. "Theory and Practice in the Choice of a Social Discount Rate for Cost-Benefit Analysis". *ERD Working Paper No. 94*. Asian Development Bank, 2007.

Appendix A: Diarrheal Disease Cost-Effectiveness, Incidents Averted per $1000 Spent



Appendix B: Example of Cost-Effectiveness Calculations

The calculations below illustrate how the assumptions and decisions discussed above are put into practice in an actual cost-effectiveness analysis, using as an example a Malawian program of conditional cash transfers targeted at adolescent girls in Malawi and evaluated by Baird et al., 2011[7]. In this evaluation, enumeration areas (EAs) were randomly assigned into a control group or one of two treatments: conditional cash transfers or unconditional cash transfers. Girls who were still enrolled in school ("baseline schoolgirls") as well as girls who had dropped out before the program began ("baseline dropouts") were eligible for the transfers. Within the conditional cash transfer (CCT) treatment group, the transfer that families received was randomly varied between $5 and $15 per month, to study the effect of differing transfer amounts on education and health outcomes. To measure potential spillover effects of the program, a randomly selected percentage (33%, 66%, or 100%) of baseline schoolgirls in each treatment EA were selected to be eligible for cash transfers. The program was run over two school years.

Below we walk through an example of calculating the cost-effectiveness of one of the two treatments in this evaluation (the CCT program) at increasing years of schooling. We list all the costs separately and model the setup of the program and the number of participants to calculate the total cost of the program, and then we divide that total cost by the total impact, which is modeled by scaling up the impact per person by the total number of participants. When an evaluation has multiple treatments, as this one does, each treatment arm requires its own cost-effectiveness calculation. Thus, the following example focuses only on the costs and

---

[7] The authors wish to gratefully acknowledge the contributions of Berk Özler, Sarah Baird, and Craig McIntosh in providing us with original cost data about the Zomba Cash Transfer Program, and in working with us to develop the cost-effectiveness model. A more complete discussion of the lessons of this cost-effectiveness calculation, including sensitivity analysis, will be available in the forthcoming J-PAL Bulletin "Improving Enrollment, Increasing Attendance."

impacts of the CCT treatment group, and the cost-effectiveness analysis for the second treatment (unconditional cash transfers—UCT) is not shown.

1. Demographics

To estimate the cost-effectiveness of the CCT program at increasing years of schooling, we first start with the basic number of students in each group so that we can later multiply the costs and impacts of the program in the correct way. We look at the 46 enumeration areas [Figure 5, Cell B11] that received the CCT program, out of the total 176 EAs [Cell B12]. Each of these EAs has an average of 16.5 baseline schoolgirls in it who are eligible for the program [Cell B13].[8] For the evaluation, a randomly selected percentage of baseline schoolgirls within each EA were chosen to participate in the program. However, we assume that all eligible girls in an EA will participate, since a scaled up program would be extremely unlikely to withhold transfers from eligible girls within the program area. (This assumption affects how we will later calculate the cost per student. For example, if there are fixed costs for serving an EA or a school, these costs will now be spread across all the students in the school.) This gives a total of 759 baseline schoolgirls [Cell B16] across the CCT EAs. Of these, 33% [Cell B15] are in secondary school and will also have their secondary school fees paid, if they meet the attendance cutoff.

Next, because we include the cost of beneficiary time to participate in the program, we estimate the time that families spend on different aspects of the program. Under the

[8] Note that we are not including impacts on baseline dropouts as they did not receive the conditionality experiment. Given that the program did have an impact on the re-enrollment of baseline dropouts, this means that the calculations produce an under-estimate of the cost-effectiveness of a population-based CCT like the one studied (rather than a commonly used school-based CCT).

demographics section, we record that the authors estimate that it will take the average parent 3.5 hours [Cell B18] to travel to and attend the initial meeting which informs them about this program. Each month in which they pick up transfers, the authors estimate that they will spend another 2.5 hours [Cell B19] traveling round trip. The program was run over 2 years [Cell B20] and the average girl met the eligibility criteria for the CCTs 7.05 [Cell B21] of the 10 months they were available. There are 3 school terms [Cell B22] each year.

2. Inflation and exchange rates

In order to bring all cost and impact calculations in terms of present value in USD for a standard year of analysis (the same year of analysis we use for all other programs in this cost-effectiveness analysis) we must use a few outside pieces of information. As we have some costs in terms of 2009 dollars, and some in terms of 2008 dollars, we will need to know the average annual inflation rates between 2009 and 2010 [Cell B24] and between 2008 and 2010 [Cell B25]. As discussed above, we use GDP deflator inflation rates for this analysis. We also note the discount rate that we will assume for this analysis [Cell B27], which is around the median discount rate used by countries employing the SOC (see Table 4), and the standard and PPP exchange rates for both the base year and the year of analysis [Cells B28-B31].

3. Enrollment and attendance impacts

Enrollment in the comparison group was approximately 79.9% [Cell J11]. Over the 6 terms of observation, the minimum transfer amount caused an additional 0.572 terms of enrollment, translating into a 9.5 percentage point [Cell J12] change each term. There was no additional impact from giving more money beyond the minimum transfer amount, and so we will use the impact estimate based on the minimum transfer amount. (This decision also affects how costs are calculated—to match the assumption about impact, we will also estimate costs as

though all girls received the minimum transfer amount.) This means that the enrollment in treatment group was on average 89.4% [Cell J13].

Attendance in the comparison group was approximately 81% [Cell J15]. Overall in 2009 (the year over which attendance was observed) the average transfer amounts caused an 8 percentage point [Cell J16] increase in attendance, conditional on being enrolled. Data are not currently available for the effect of the minimum transfer amount on student attendance (the sample size is too small to precisely measure effects among this group) so we assume that the minimum and average transfer sizes have the same effect on attendance. This means that the conditional attendance in the treatment group was on average 89% [Cell J17].

4. Opportunity cost of parent time

To estimate the costs that families faced in participating in this program, we need to know the opportunity cost of their time, in addition to how much time they spent participating, which was recorded above. This evaluation did not gather information on average daily wages in the study area, nor were we able to find district-level data on daily wages. Using information from Jessica Goldberg's paper 'Kwacha gonna do?' we estimate a rural Malawian daily wage of 87.5 MWK [Cell J20], or approximately 62 cents [Cell J21]. This translates into an hourly opportunity cost of parent time of around 8 cents [Cell J22].

5. Administration costs

We also need to estimate how much was spent on the administration of the program. Because the CCT treatment was implemented alongside the UCT treatment, administrative costs were only available in aggregate, and so we had to estimate how much of those total administrative costs were attributable to the CCT arm. The authors estimated that the total administration costs in year 2 were approximately $80,000, split between $10,000 of fixed costs

[Cell J26] and $70,000 of variable costs [Cell J25]. Approximately 33% [Cell J27] of these variable costs were spent on the distribution of transfers in all treatment EAs, while 67% [Cell J28] were spent on gathering attendance data to verify conditionality in the CCT EAs.

It should be noted that these administration costs covered both baseline schoolgirls and baseline dropouts, so if we are restricting our estimate of the impacts to baseline schoolgirls, we must include only the costs of administration needed to include the baseline schoolgirls in the program. And when we calculate the costs for "the CCT program", we need to exclude the costs that were spent on administration on the other treatment, unconditional cash transfers. This can be done by assuming that the cost of transfer distribution was the same across EAs, and so the cost of transfer distribution within one treatment arm was simply equal to (for instance) the number of CCT EAs as a proportion of all of the treatment EAs (46/46+27) = 63%. Based on these considerations, we separate out three types of administrative costs:

- **Administrative costs of transfer distribution**, spent just on baseline schoolgirls in CCT EAs [Cell J29] = $70,000 * 33% of administrative costs spent on distribution * 63% of distribution costs spent in CCT EAs

- **Administrative costs of verifying attendance**, spent just on baseline schoolgirls in CCT EAs [Cell J30] = $70,000 * 67% of administrative costs spent on verification * 87% of verification done for baseline schoolgirls

- **Fixed administrative costs**, proportion attributable to CCT EAs [Cell J31]
= $10,000 * 63% of fixed administrative costs spent in CCT EAs

These costs, as well as others based on the demographics and program characteristics outlined above, are brought together in the ingredients section below.

6. Ingredients section

The information above provides all of the necessary details on the costs of this program, but it is necessary to make sure that all of these costs are expressed in the same currency, from the perspective of the same year, and taking into account the present value of the cost stream (discounting).

a. Listing Ingredients: The total cost of the initial census for all EAs (CCT, UCT, and control) was $100,000 [Cell J32]. We have distributed this proportionally to each EA by multiplying $100,000 times 26.1% (46/176) and included the fraction of the initial census for the CCT arm of the program [Row 35]. We then bring in the administrative costs of transfer distribution, gathering attendance, and the fixed administrative costs in Rows 36-38, and the opportunity cost of parents' time to attend informational sessions in Row 39. As we are modeling the cost-effectiveness of the minimum transfer amount, we assume that each girl gets a total of $5/month for all months for which transfers are given [Rows 40-41]. For each eligible girl in secondary school, the program must pay 3,000 MWK per term of the school year to pay for her secondary school fees, included in Row 42. Lastly, we include the opportunity cost of parents' time to pick up transfers each month in Row 43.

   Next, in moving from left to right in the spreadsheet, we follow a number of steps to bring all the costs that we have now listed into a common currency and a common year, and discounted to reflect the present value stream of costs. This same order of steps is used for all the programs included in our larger cost-effectiveness analysis on student participation.

b. Currency Exchange: The only ingredient whose cost is not reported in USD is the secondary school fees for girls [Row 42], therefore this must be converted using the

exchange rate noted above in Cell B28. Now, all costs are expressed in terms of the same country's currency [Column G].

c. Deflation to Base Year: The opportunity cost of parents' time was available only in terms of 2009 currency, and so it is necessary to deflate this back to the base year (2008) using the inflation rate noted above in Cell B24. Now, all costs are expressed in terms of the same country's currency, in the same year [Column H].

d. Present Value of Cost Streams: This was a two-year program, and many of the ingredients must be purchased over both years of the program. For all ingredients for which this is true (everything except Rows 35 and 39), it is necessary to calculate the present value of the cost stream using the discount rate identified above in the Demographics section[Cell B27]. Now, all costs are expressed as the present value of the cost stream in terms of the same country's currency, in the same year [Column I].

e. Inflation to Year of Analysis: The costs in Column I are still expressed in terms of the present value in the base year of the actual program (2008), and so it is necessary to inflate them forward to the year of analysis (2010) using the inflation rate noted above in Cell B25. Now, all costs are expressed as the present value of the cost stream in terms of the same country's currency, in the year of analysis [Column J].

The total cost of the program can now be calculated as the sum of all of these present value streams [See the "Total Cost" cell in the "Cost-effectiveness box below: Cell B55].

7. Aggregating impacts

We also need to calculate the total impact for the entire group for which we just totaled the costs. To calculate total impacts, we calculate the total years of schooling achieved in the treatment [Row 57] and comparison [Row 56] EAs, and subtract. We calculate the total years of

schooling achieved by multiplying the enrollment rate [Cell C47] times the conditional attendance rate [Cell D47] and then multiplying by the number of eligible girls [Cell B47]. Note that we do not use the actual number of baseline schoolgirls in comparison EAs to calculate the total impacts in the "comparison group"—this is because we are not calculating the total years of education that would occur in the 88 comparison EAs, but rather the total years of education that would occur if the 46 treatment EAs has not experienced the program. So, we take the level of enrollment and attendance experienced in the comparison group, and then scale that up by the number of girls in the treatment we are analyzing. We estimate that girls in treatment EAs experience a total of 113 additional years of schooling over one year of the program [Cell F50], and at a 10% discount rate this works out to be 216 additional years of schooling over the two-year life of the program [Cell F51].

8. Cost effectiveness

The final cost per additional year of schooling [Cell B58] is calculated as the total cost [Cell B55] divided by the total impact over the life of the program [Cell F51].

Appendix C: Worksheet for Gathering Costs

Framework for Gathering Cost Data

1. Study Information	
Study Title	
Authors	
Location	
Base Year (year in which program launched)	

2. Demographics	Number	Unit	
Intervention Duration			
Number of Individuals/Schools/Communities in T			
Number of Individuals per School/Class/Community			
Exchange Rate Used (if costs already in USD)			
Local Daily Wage (i.e. opportunity cost of user time)			

3. Cost Data

Please bring together all of the costs of running the project you're evaluating (distinct from the costs of actually evaluating it) and fill out information on all cost items that apply to this program. The goal of this information is to get an idea of how much it would cost a government or NGO to replicate the program, so it's important to consider whether the project had any cost categories covered for you (i.e. by piggybacking on existing NGO infrastructure) which a scaled-up model of your program would have to pay for. One useful way to think about this cost gathering exercise is like writing a recipe for the intervention you are testing, and you need to come up with a list of all ingredients.

~What is every single ingredient necessary for this program to have the observed impact?

~How much of each ingredient is needed?

~How much does one unit of this ingredient cost?

~When is this ingredient used?

The categories listed are intended as ideas for the kinds of costs you might incur. **Please add your own categories and items as necessary.**

Exchange → Deflation → Present Value → Inflation

Costs of Administration & Targeting	Unit Cost	Currency (Location & Yr)	Units Req'd (per year)	Years Required	Total Cost/Yr, Local Currency	Total Cost/Yr, Yr Incurred USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
Hourly wages, surveyors for targeting survey?									
Transportation per diem for surveyors?									
Hourly wages, data entry staff?									
[Insert your own cost items]									
[Insert your own cost items]									

Exchange → Deflation → Present Value → Inflation

Costs of Marketing & Education	Unit Cost	Currency (Location & Yr)	Units Req'd (per year)	Years Required	Total Cost/Yr, Local Currency	Total Cost/Yr, Yr Incurred USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
Development of outreach leaflets?									
Printing of outreach leaflets?									
Hourly wages of trainers, for marketing training?									
Per diems of trainers, for marketing training?									
Transportation per diem, for marketers?									
Hourly wages of marketers?									
[Insert your own cost items]									
[Insert your own cost items]									

Exchange → Deflation → Present Value → Inflation

Cost of Materials & Productive Assets	Unit Cost	Currency (Location & Yr)	Units Req'd (per year)	Years Required	Total Cost/Yr, Local Currency	Total Cost/Yr, Yr Incurred USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
Procuring productive assets?									
Productive asset (per client)?									
Transportation fee (per asset)?									
Cash subsidy (per client per month)?									
[Insert your own cost items]									

[Insert your own cost items]									

Exchange → Deflation → Present Value → Inflation

Costs to Clients	Unit Cost	Currency (Location & Yr)	Units Req'd (per year)	Years Required	Total Cost/Yr, Local Currency	Total Cost/Yr, Yr Incurred USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
Hourly wage of clients incorporating new technology?									
Subsidies for clients (negative cost of cash subsidy)?									
[Insert your own cost items]									
[Insert your own cost items]									

Exchange → Deflation → Present Value → Inflation

[Insert Your Own Category Here]	Unit Cost	Currency (Location & Yr)	Units Req'd (per year)	Years Required	Total Cost/Yr, Local Currency	Total Cost/Yr, Yr Incurred USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
[Insert your own cost items]									
[Insert your own cost items]									