

2 Making Aid Work

Abhijit V. Banerjee and Ruimin He

One would think that giving away money ought to be easy. After all, there are so many who need it so badly, and we have a pretty good idea of where they are. Yet, rather remarkably, we seem to have arrived at a point where more or less everyone agrees that aid giving is not working the way it ought to.

Part of the problem is that most people are not actually giving away money—or rather, while they often give away money, the ultimate recipients do not get it as a straight gift of money. Somewhere along the chain, a non-governmental organization (NGO) or a government is responsible for turning this money into schools, or hospitals, or roads, or whatever else the people are supposed to get.

2.1 What Makes Giving Away Money Hard?

The bias against just giving people money stems in part from the feeling that the best use of the money may not be to spend it on consumption. This is plausible, and indeed likely: there is some relatively convincing evidence that many people do not invest as much in their businesses and their children's human capital as the rewards to such investment seem to warrant. What is less clear is why people could not be relied on to make the right investments on their own—in which case it would be enough to hand them the money. One possible reason may have to do with the lack of self-control. It may be too tempting, especially for poor people, to spend the money on something they need right away. The incompleteness of the intrafamily contract is another reason not to trust the family with the money: parents may put too little weight on improving their children's earning capacity because they do not expect to share in their children's prosperity. And, of course, people may not know what is good for them.

There was a time when many of these kinds of arguments could not be made among respectable economists. They were seen as a transgression against the

freedom of the individual, and bad economics to boot. Now the pendulum swung in the other direction: it is an item of faith in the development community that no one should be giving away money. It is not clear what, if any, evidence lies behind this shared conviction. Certainly no one has done an experiment of showering large gifts of money on poor people in poor countries and then following them to see what they do with the money and what happens afterward.

A very different kind of argument for giving people goods rather than money comes from the fear that if you are seen handing out money, even those who have enough of their own may want to pretend to be needy. The advantage of providing public services rather than money is that the nonpoor may not value them enough to make it worth their while to simulate poverty. The very rich in the United States, after all, choose to pay for their children to go to private schools, even though their children are entitled to go to public schools at no charge, simply because they feel that the public education system is not good enough up to the standards to which they aspire. This saves the government the trouble of finding teachers for these extra children. But these parents do not hesitate to claim any tax deductions that they may be entitled to, which suggests that if the government was giving away money instead of schooling, the rich would be in the queue with everyone else.

It is, however, not at all obvious that one cannot give away money without opening the floodgates. After all, the rich value their time: making it necessary to queue up in order to collect the money should discourage those who do not need it.

The broader point here is not to deny that giving away money has significant disadvantages, but to emphasize that we know very little about how serious these disadvantages might be. In particular, are these costs necessarily large enough to outweigh the significant costs of trying to give away anything other than money?

2.2 Delivering Goods and Services to the Poor

If you do not want to give away money but still think it is worth trying to help the poor, you would have to give them things: roads, schools, banks, hospitals, fertilizer. Giving away things is more work than giving away money for a simple reason that someone needs to produce them: roads have to be built, teachers hired and trained, fertilizer produced, and hospitals kept in good repair. It would be simple if it were just a matter of paying for the roads, fertilizer, and the other things, but that is just the first step. Then we would have to make sure that the roads are built to the required standards, the teachers te-

the fertilizer gets to the right people, the process of delivery is not subverted by corruption or bureaucratic incompetence.

What is perhaps even more difficult is to be reasonably confident that the money is being spent on things that are really worth getting, in the usual sense of being at least as good as any other way of spending the same amount of money. After all, to take one popular example, there are many ways to spend money on promoting education: build more buildings, hire more teachers, provide free textbooks, distribute free uniforms, put flip charts in classrooms, set up computer labs in every school, provide a bonus for teachers who teach well, serve hot meals in school, and much more. Indeed every one of the interventions listed here (and many more) has been tried somewhere in the world in the past few years. This diversity in part reflects differing needs, but often it is just ignorance. Lacking a clear sense of what works, well-meaning donors will choose what their intuition suggests, even though it may be very different from what the donor next door believes. Both seem to believe that they are right.

There is also no guarantee that it makes sense to spend the money on directly promoting education. The fact that people are not getting an education on their own might reflect the lack of jobs for those who have an education. Pushing out more graduates who will not find jobs either may actually be counterproductive, because it might reinforce the lack of faith in the value of education. The best way to promote education may be to create jobs. Or we may even want to look beyond education. Perhaps one should invest in health and leave education to private initiative—or forget about both and go for fertilizers. How should we decide where to go?

It is no surprise that the process of helping the poor by giving them access to goods and services is fraught with difficulty. There needs to be a system for picking the right project and a system for making sure that the project is carried out as it should be, and for figuring out how much people are getting out of it and whether it continues to be what they need or want.

For this, donors need to get involved in the process of decision making and delivery at the ground level, though the exact nature of the involvement can vary substantially. The actual production of the good is usually contracted out, though even this is not always the case. Both the process of delivery and its impact have to be assessed, though once again there is a choice between doing it yourself and contracting it out. And the broader strategy needs to be worked out, based, one presumes, on knowledge of the situation on the ground. This might mean consulting local experts. Even carrying out some new research as a prelude to the intervention is not out of the question.

In all of this, the donor will typically work with one or more local organizations, be it government departments or NGOs. Some of these local partners

may have their own sources of funding, which will allow the donor to leverage its resources.

Whatever the exact strategy, whether the donor does the monitoring or preliminary research, or whether these are contracted out, they involve substantial expenditures over and above the direct cost of delivering the good (or service) to the ultimate beneficiaries. To get a sense of how large these expenses might be, we note that between the years 1996 and 2001, the World Bank administrative budget averaged US\$1,401 million per year (World Bank 2001, appendix 5; World Bank 2003b, appendix 1), while the total World Bank International Bank for Reconstruction and Development (IBRD) and International Development Association (IDA) loans per year (World Bank 2000b, 10, World Bank 2002b, 26) averaged US\$15,615 million and US\$6,154 million respectively. The Operations Evaluation Department of the World Bank calculates ex post economic rate of returns for the projects that it evaluates (World Bank 2002a, statistical appendix table 12), but only for certain sectors.¹ It finds that the median revised (ex post) economic rate of return for both the IBRD/IDA/blend lending operations for fiscal years 1996–2001 exits (the date the project leaves the World Bank's active portfolios) was about 20 percent. The aggregate economic return on the bank's portfolio was 20 percent of \$15.615 billion plus \$6.154 billion, or about \$4.400 billion. It follows that the benefits from World Bank lending are lower by almost a third because of administrative costs.

As we will discuss, we should not expect to entirely avoid these costs by switching to donating money rather than goods. For now, however, the relevant point is that these costs are large.

2.3 Evaluating Donor Effectiveness

Given everything that they spend on the design and management of aid projects, are donors getting what they hoped for? The short answer is that we do not know. Part of the problem is data: As table 2.A.1 in the appendix shows, most of the larger public donor organizations (as opposed to private foundations) do evaluate their projects, but they usually stop short of a summary quantitative assessment of the social impact of the project, such as a rate of return. Of the donors listed in this appendix, only the World Bank reports rates of returns, and then only for certain sectors; education, health, and nutrition, for example, are left out. This reflects, in part, doubts about whether it makes sense to try to reduce the many dimensions of a project outcome to a single rate of return. In part, it also reflects the inherent difficulties of coming up with

rate of return: How do you come up with the right counterfactual that tells you what would have happened in the absence of the project?

What some of these donors do instead is to assess overall project performance combining both process evaluation (“the right number of schools were built”) and impact evaluation (“the children’s test scores improved”) on a set scale such as satisfactory/unsatisfactory or successful/unsuccessful (table 2.A.1).² However, this evaluation is often carried out by those who are also involved in the implementation of the program, making it somewhat hard to know what to make of the results. Finally, many of these organizations do not allow the public access to their assessments. Of the eight organizations listed in table 2.A.1, only the World Bank and the Asian Development Bank (ADB) have their projects assessed by a formally independent organization, report project level assessments on a set scale, and allow the public access to the assessment results.³ These are the only two, therefore, that offer the possibility of delving deeper.

This is far from ideal. The World Bank is not just any donor organization: it is probably the most visible organization of this class, with all the constraints that come from being in the public eye. It is formally responsible to those who provide its financing, which are the governments of a handful of rich countries. Moreover, it gives only loans (albeit on very attractive terms). It is also an organization that attracts and employs many of the best minds thinking about development today. Perhaps most important, it sees itself as a leader in the efforts to promote development in the world. This probably means that its projects need to be evaluated not just in terms of what they directly achieve but also in terms of how they shape efforts outside the bank to promote development. In a previous paper, we (Banerjee and He 2003) try to evaluate the bank’s achievement as a leader and conclude that there is no evidence that others are following its lead. But we also argue that the bank is ideally placed to take this leadership role and that it is important that it does so, which obviously implies that we must take it seriously in evaluating the bank’s performance.

The ADB is also quite special, being the one multilateral funding organization that has close ties with the Japanese government. We may expect it to have been influenced by the Japanese government’s rather distinct view of economic policy.

Given that the World Bank and ADB evaluations are on different scales and the evaluators have potentially different standards, there is no point in trying to compare these two organizations directly. One could, of course, take their assessments at face value. In 2002, the Operations Evaluation Department (OED) of the World Bank wrote that “at the project level, the outcomes of Bank-financed projects continue to improve,” and more than 60 percent of all

projects evaluated each year since 1990 have had satisfactory outcomes (World Bank 2003a, xii); while the ADB wrote that the proportion of successful projects or programs (by year of completion) has been more than 50 percent since 1997 and has been trending upward since 1989 (Asian Development Bank 2003, 37).

The problem is that we do not really know what to make of the scale of use. What does it, at some absolute level, mean to say that the World Bank or the OED feels that the project was satisfactory? How much of this assessment reflects, for example, what they expected (which we do not know) rather than some objective that we all share? Clearly there are many who would join Easterly in his reluctant conclusion that the last fifty years of aid giving by the World Bank (which for many years employed Easterly) has achieved relatively little (Easterly 2001, 2003). It seems safer, therefore, not to put too much weight on how the OED (or its equivalent elsewhere) feels about the project and to focus on how assessments vary across different projects or sets of projects for the same organization.

World Bank or ADB projects vary considerably in the degree of donor involvement, as measured by the share of the project financed by these organizations. If we assume that putting more money into the project reflects greater commitment to the cause, we can use this ratio as a measure of the bank's priorities.

What can we say about the World Bank's priorities? For each project approved (i.e., launched) between 1994 and 2001, the World Bank reports the share of World Bank funding in total funding for that project.⁴ For the period 1987–2001, we also have the evaluation of projects by sector,⁵ averaged over three-year periods (1987–1990, 1990–1993, 1994–1997, and 1998–2001).⁶ We label these four periods 1, 2, 3, and 4. We then regress the share of World Bank funding in a particular project approved in period t on the average evaluation in period $t - 2$ of the sector that it belongs to [*Prevperf*], and the improvement of its evaluation between $t - 1$ and $t - 2$ [*Diffperf*]. We control for fixed differences across sectors, countries, and periods and cluster errors by sector.

The results in column 1 of table 2.1 show that when a sector's performance improves, projects in that sector get a higher fraction of their financing from the bank (the *DiffPerf* coefficient is positive). But it also helps to start at a low base (conditional on the same degree of improvement, projects in sectors that started with a worse record get more money: the coefficient on *Prevperf* is negative), which immediately implies that if two sectors have shown the same improvement, the one that is doing worse will get more money from the bank.

A similar pattern emerges when we look at the total amount of money allocated to each sector. For four sectors—agriculture, finance, technical a:

Table 2.1
OLS regressions on project selection

	World Bank (1)	ADB (including multisector/others) (2)	ADB (excluding multisector/others) (3)
<i>PrevPerf</i>	-.002553* [.0011]	-.006767* [.0033]	.0004253 [.0047]
<i>DiffPerf</i>	.001852* [.00051]	-.004361* [.0014]	-.001184 [.0020]
<i>N</i>	1513	519	468
Adjusted <i>R</i> ²	.29	.33	.39

Note: Dependent variable: percentage of individual project that is funded by the World Bank. Significant at the 5 percent level of significance. Robust standard errors reported.

Table 2.2
Cross-sectional panel regressions on project selection

	World Bank (1)	ADB (2)
$\text{Log}(\text{expenditure}_{t-1})$.7677* [.028]	.1523 [.18]
Outcome_{t-1}	-.002256 [.0026]	.000925 [.0089]
$\text{Outcome}_t - \text{outcome}_{t-1}$.001594* [.00075]	-.0004837 [.0073]
Number of observations	16	45
Number of groups	4	7

Note: Dependent variable: $\text{Log}(\text{expenditure}_t)$.

* Significant at the 5 percent level of significance. Robust standard errors reported.

tance, and water and sanitation⁷—the World Bank provides data from 1974, when data start, to 1993⁸ for every block of three years (with the exception of 1992–1993) on the sector outcome measured by OED evaluation,⁹ and the log of total expenditure on that sector.¹⁰ Using this data set, we regress using the Arellano-Bond linear dynamic panel data estimator [$\text{log}(\text{expenditure}_t)$], on [$\text{log}(\text{expenditure}_{t-1})$], [outcome_{t-1}], and [$\text{outcome}_t - \text{outcome}_{t-1}$], correcting for period and sector effects. The results are shown in column 1 of table 2.2. They show that increased spending for a particular sector is associated with an improving trend in sector performance over the immediate past. Once again, having started from a lower initial level of performance helps, but the coefficient is not significant in this case.¹¹

When we do the same exercise for the ADB, we get very different results. Both past performance and improvement in performance seems to have a negative

impact on the allocation of its funding (column 2 of table 2.1). However, the result is very sensitive to the inclusion of the "Multisector/Others" sector where there were only two evaluations between 1986 and 1989 (both successful). Once this sector is dropped, the past level and the improvement in the become insignificant (column 3 of table 2.1). All that matters then are cross-country differences and perhaps cross-sector differences.¹²

The results for the ADB seem consistent with the view that its priorities are largely set by high-level decisions and are not particularly subject to any short-term influences. This may be a good thing because it insulates the organization against the influence of fads and internal political shifts, but it clearly also prevents the organization from learning from its experience.

The case of the World Bank is more complicated. It clearly does not give priority to sectors that have been performing the best over the immediate past, which is what, under the (possibly brave) assumption that past performance is a reasonable index of what we might expect in the immediate future, we would have been the way to maximize immediate impact. But it does favor the sectors that have been improving the fastest. One way to rationalize this may be to assume that the World Bank sees itself as a leader in the development community. As a leader, it would make sense for it to try to promote those sectors where the potential for improvement is the highest rather than those where the current record, the possibility of success is the highest. Sectors that have been improving fast over the past few years, but still have some distance to go, therefore become exactly the sectors the bank would want to favor.

It is, however, possible to take a more cynical view of the same evidence. In this view, the bank is excessively influenced by shifts in current fashions in development thinking. The reason, in this view, that we see the bank reacting to improvements is that these improvements shift fashions. When something is done unexpectedly well, it is easy to get excited about it, even if, on balance, it is doing worse than better-established options.

The question, in the end, comes down to whether the projects that are given priority are doing what they were intended to do. One way of looking at this is to examine the correlation between the fraction of planned project financing that was to come from the bank and the performance of the project according to the bank's evaluators, after controlling for fixed differences across sectors and countries, the length of the project, the year when it was approved, and the year of the evaluation. The results are shown in column 1 of table 2.1. As we have already reported (Banerjee and He 2003), there is a negative and significant correlation between the priority that the bank originally gave to a project (measured by the fraction of financing that was supposed to come from the bank) and its performance. Bank-favored projects seem to do worse from an ex ante point of view relative to other projects in the same sector.

Table 2.3
On effectiveness of fund allocation

	World Bank (planned) (1)	World Bank (actual) (2)	ADB (planned) (3)	ADB (actual) (4)
% Funding	-.123** [.038]	.064 [.090]	-.585 [.54]	-.837* [.49]
Length	.0054 [.016]	.0100 [.016]	-.0705* [.038]	-.0722** [.035]
<i>n</i>	664	664	137	136
Adjusted <i>R</i> ²	.26	.24	.18	.20

Note: Dependent variable: Outcome of individual projects. Regressions include sector dummies, country dummies, year of approval dummies, and year of closing dummies.

**Significant at the 5 percent level of significance. Robust standard errors reported.

*Significant at the 10 percent level of significance. Robust standard error reported.

This negative relation goes away if we replaced the share of planned cost that the bank was supposed to pay for with the share of actual cost (column 2 of table 2.3). Basically, if projects are going really badly, the bank cancels its promised contribution to them. But even with the help of this corrective procedure, the correlation between performance and funding is nowhere near being positive and significant. Being a bank priority does not help you perform better, even after the cancellations.

We repeat this exercise for the ADB, using projects evaluated between 1997 and 2002.¹³ Here percentage funding is defined as the [Loan amount approved]/[Expected project cost] in column 3 of table 2.3, and [Loan amount disbursed]/[Actual project cost] in column 4 of table 2.3. We find that the outcome rating is not significantly affected by the loan amount approved, but the amount disbursed has an effect that is negative and significant at the 10 percent level.

The lack of a positive correlation between funding and performance, in the case of the ADB, seems unsurprising, given everything else we have seen about the way they target (or rather, do not target). In the case of the World Bank, these results are consistent with the view that the bank is faddish. They can also be explained by assuming that the bank is particularly ineffective at running its projects.

But the results are also quite consistent with any view that has the World Bank playing the role of a leader and prioritizing projects that others, more focused on immediate impact, would not choose. After all, we already knew that they have not given priority to the sectors that had performed best in the past. All that this evidence really shows is that this is true within sectors as well. Of course, it is not clear that the World Bank is particularly effective as a leader,

and to the extent that we take this as given, we may still want it to focus on its rather limited direct impact.

In the end, there is little that is reliable that we can say about donor performance. The most we can say is that we found no prima facie evidence of effectiveness.

2.4 What Limits Effectiveness?

Donor organizations are in many ways very much like other organizations, they share many of the standard organizational constraints. Organizations like the World Bank obviously need many people to act and take decisions on their behalf, and there is nothing to guarantee that they have the right incentives; in particular, in an organization that lives by doing projects and making loans one achieves prominence by rejecting projects and refusing loans. For this reason alone, most people on the implementation side of the bank, in either its departments or country missions, are probably somewhat biased in favor of making something happen. Add to this the fact that they are the ones who deal with the potential recipients, and we have a recipe for a degree of enthusiasm and irresponsible lending.

There is also obviously the possibility of ideological conflict within the organization. We recently saw a public example of such a conflict in the World Bank that ultimately led to the resignation of the person in charge of the 2000–2001 *World Development Report*. There must also be other fights that have less to do with ideology than with personalities and individual ambitions. All those involved in these fights must be tempted to use the power to sanction projects to help their supporters and punish their enemies.

There is also the fact that being a donor organization involved in development makes one someone that many people want to influence—ranging from the U.S. government to NGOs hostile to the U.S. government. These pressures are probably easier to resist for an organization that has an explicit ideology or an acknowledged political master, like the Salvation Army or the U.S. Agency for International Development, than a nominally apolitical organization like the United Nations Development Program (UNDP) or the bank. For such organizations, the challenge is often in maintaining their reputation for being open to different views while continuing to make the right choices about the allocation of funds. Resisting the temptation to placate the different sides by conceding to some of their less merited demands must be a challenge for these organizations.

One of the biggest problems, and one that is discussed all too rarely, is the lack of an explicit scientific basis for their decision making. An eloquent example is Lant Pritchett, a long-term bank employee:¹⁴

Nearly all World Bank discussions of policies and project design had the character of “ignorant armies clashing by the night”—there was heated debate amongst advocates of various activities but rarely any firm evidence presented and considered about the likely impact of the proposed actions. Certainly in my experience there was never any definitive evidence that would inform decisions of funding one broad set of activities versus another (e.g., basic education versus roads versus vaccinations versus macroeconomic reform) or even funding one instrument versus another (e.g., vaccinations versus public education about hygiene to improve health, textbook reform versus teacher training to improve health, textbook reform versus teacher training to improve educational quality). How can this combination of brilliant well-meaning people and ignorant organization be a stable equilibrium? (Pritchett 2002, 251)

A World Bank publication from a few years ago, *Empowerment and Poverty Reduction: A Sourcebook* (Narayanan 2000), provides an excellent case study of the kinds of policies that result from these deliberations. The *Sourcebook* is meant to be a catalogue of what, according to the bank, are the right strategies for poverty reduction. These are also, we presume, strategies into which the bank is prepared to put its money. It provides a long list of recommended projects, which include computer kiosks in villages, cell phones for rent in rural areas, scholarships targeted toward girls who go to secondary school, schooling voucher programs for poor children, joint forest management programs, water users’ groups, citizen report cards for public services, participatory poverty assessments, Internet access for tiny firms, land titling, legal reform, microcredit based on group lending, and many others.

While many of these are surely good ideas, the book does not reveal how we know that they work. We now know that figuring out what works is not easy. There is a large literature documenting the many pitfalls of the usual intuitive approach to program evaluation. When we do something and things look as if they are getting better, it is tempting to think that it was all because of what we did. The problem is that we have no way of knowing what would have happened in the absence of the intervention. The simplest and best way to avoid this problem is to do a randomized evaluation where we assign the intervention to a randomly selected subset of the set of potential locations and compare those who got it with those who did not. This mimics the procedures used in trials of new drugs, which is the one place where, for obvious reasons, a lot of care has gone into making sure that only the things that really work are approved. In many ways, social programs are very much like drugs because they have the potential of transforming the life prospects of people. It seems appropriate that they should be held to the same high standard.¹⁵

Of course, even randomized trials are not perfect. Something that may work in India may fail in Indonesia. Ideally there should be multiple randomized trials in varying locations. There is also no substitute for thinking. There are often

good and clear reasons why what works in Kenya will not work in Cameroon. And there are times when randomized experiments are not feasible (more than that later). However, with all that, it is hard to imagine a good reason for spending a lot of money without having done at least one successful randomized trial, assuming that a randomized trial is possible. When we talk of hard evidence, we will therefore have in mind evidence from a randomized experiment or, failing that, evidence from a true natural experiment, by which I mean an accident of history that created a setting that mimics a randomized trial.¹⁶

What is striking about the list of strategies offered by the *Sourcebook* is the lack of any distinction between strategies that can claim to be based on hard evidence and the rest. In fact, to the best of our knowledge, only one of the strategies—schooling vouchers for poor students in Colombia—has been subjected to a randomized evaluation, and that was because it was politically necessary to allocate the vouchers by lottery. Comparing those who won the lottery with those who did not provided the perfect experiment for studying the impact of the program, and the study by Angrist et al. (2002) takes advantage of it. Yet the results from this study receive no more weight than any of the other programs.

Indeed most of these programs are recommended on the basis of very hard evidence. Legal reform, for example, is justified by asserting that “the extent to which a society is law-bound affects its national income as well as its level of literacy and infant mortality.” This may be true, but the available evidence, which comes from comparing the more law-abiding countries with the rest, confounds too many things to warrant such a confident recommendation.

And some programs, it seems, no amount of negative evidence can stop. A favorite example is the *Gyandoot* program in Madhya Pradesh in India, which provides computer kiosks in rural areas. The *Sourcebook* acknowledges that this project was hit hard by lack of electricity and poor connectivity and “currently only a few of the Kiosks have proved to be commercially viable.” It then goes on to say, entirely without irony, “Following the success of the initiative” (p. 80).

2.5 Why Do People Resist Evidence-Based Policymaking?

Lant Pritchett (2001) goes on to argue that the resistance to hard evidence is in part a reflection of the mixed motives of those who give and receive aid. Even where there is no real corruption, as in the bank, the problem is that many of these people are true believers and see no intrinsic value in rigorously testing the policies that they are advocating. Although they recognize that good

dence might help them win friends, they also worry that it might work against them. Someone might misread the evidence, or, as chance will have it, the evidence may just refuse to cooperate. Hard evidence is simply not worth the trouble, especially if eloquence and a few carefully chosen examples can carry the day.

We do not doubt that this is a piece of what goes on, but it explains the motives of only those who have things going their way. But the bank, for example, is a contentious place. For every one person who likes where policy is currently headed, there are probably at least two who would like the tides to turn. These people have a strong incentive to look for hard evidence, since there is no other way they can upset the status quo, and it is hard to believe that they could not do a proper test of the intervention if they really wanted to. In other words, if the advocates do not provide the necessary hard evidence, we would expect their opponents to do so.

The fact that we see very little of this kind of competition by evidence suggests to us that the deeper problem is not strategic resistance to evidence, but rather a view shared by most people in the development community that basing policy on hard evidence is simply not practical. This is consistent with our experience in talking to senior officials in donor organizations, who seem to genuinely believe that there is no real alternative to the current system of decision making.

Their objections to the idea that policy should be based on evidence typically fall into one of two categories. First, there is the fear that requiring that every initiative be justified in terms of hard evidence will bias decisions in favor of what is measurable and easy to evaluate. Second, there is the conviction that at this point, there is so little that can be justified in terms of randomized trials that to rely exclusively on this evidence is tantamount to considered inaction.¹⁷

2.6 The Feasibility of Evidence-Based Development Policy

We feel that both of the concerns articulated in the previous paragraph are substantially exaggerated. We are certainly not saying that every policy action needs to be justified in terms of hard evidence. There are things like macro-policy that are very hard to evaluate properly. The problem is that once something is big enough ("currency boards," "democracy"), there is going to be no way to know what would have happened in its absence. And yet there are clear examples of policies that, most people would agree, make very little sense ("overvalued fixed exchange rates," "a pension plan that is headed for bankruptcy," for example). There is no question that helping governments in their efforts to get out of these indefensible policy positions is a good use of donor money.

On the other hand, there are many macro interventions that do allow for limited micro evaluation. For example, while decentralizing political power as a macro reform, we could learn a lot about it by looking at the impact of an initial pilot, where the reform is implemented only for certain areas, chosen randomly from a larger set. Where this is not possible and the decision has to be based purely on theoretical reasoning, evidence from micro studies can be useful because it can help us choose the right theory. Obviously how far a donor is prepared to go down this evidence “quality ladder” will depend on the donor. What is key is that she has a sense of what she is giving up—the possibility that there may be other projects for which we have much more reliable (and reassuring) evidence.

The other side of this same concern is that requiring evidence discrimination against projects that promote less measurable outcomes, such as female empowerment. It is true that historically the focus of economic measurement has been on concepts like consumption and income rather than empowerment, but when Chattopadhyay and Duflo (2001) needed a measure of female empowerment in the context of public action in Indian villages, they used the fraction of questions asked by women in village meetings. While this is not perfect, it is not obviously worse than using income to measure well-being, which we regularly do. We are therefore optimistic that once we commit ourselves to a measurement, the interaction of the donors and the evaluators will generate a range of good measures of most things that are relevant.

To address the other main concern, that basing action on evidence will lead to paralysis, we carried out a crude but useful exercise. We began by searching for interventions that, at the time when the piece was written (2004), had been subject to an evaluation based on random assignment (though not necessarily as a part of an experiment) and appeared to work. To come up with this list we asked researchers in the Bureau for Research in Economic Analysis of Development (BREAD) for references and used summary papers by Krashinsky (2003), Behrman and Knowles (2003), De Cock et al. (2000), and World Bank Group 5 of the Commission of Macroeconomics and Health (2001) as starting points for a literature search. In addition, a Web-based search was used. In these searches we deliberately left out regulations, such as tobacco taxes and bans on tobacco. The table of interventions in the appendix (table 2.A.2) lists all the papers that were eventually included in our list. It is meant to cover every category of micro intervention that we would find that has been subjected to evaluation.

From these papers, we highlighted the subset of programs that, based on the currently available evidence, look sufficiently good that it would be worth implementing them on a global scale. This cut was based on three criteria: a program must be sustainable without a strong intervention by the

searchers;¹⁸ the evidence must come from a run of the program where it was randomly placed; and finally, in that randomized trial, it must have had a significant positive impact on at least one of the initially chosen objectives. If two experiments showed different results, wherever there was a clear methodological difference, we favored the one that has the better experimental design. We also favored experiments that showed improvements in outcomes that have direct economic relevance—education, incidence of disease, weight—over outcomes that are only potentially correlated with economic outcomes—the presence of antibodies, for example. Finally, we favored experimental research by economists over experiments by clinical researchers, on the grounds that economists are more sensitive to problems relating to delivery.¹⁹ In the end, there were few cases where we had to exercise any judgment.²⁰

We left out all programs that simply gave away money but included programs (like school vouchers) that give people money that can be used only for a specific purpose. Finally, we were not quite sure of how to deal with a program like PROGRESA in Mexico, which makes an income transfer to mothers who send their children to school. The problem is that we have no idea of how much we would need to pay to get children into schools and how much of what was paid was a pure gift. We therefore treated it as a separate category.

The goal here is purely illustrative: We want to demonstrate that one can come up with a long list of interventions that have been shown to work based on a randomized evaluation. We are not at all suggesting that the interventions listed are the only ones that work or even that they are the most efficacious among those that do work. Nor do we suggest that the programs that we list but decided not to scale up are necessarily worthless. It is entirely possible (indeed, we hope that it is true) that there are many other interventions that do work but are not included in our final list because we could not find a randomized evaluation applying to them. Other programs that were left out because there was a high degree of intervention by the researchers might have also worked with less intervention. We simply do not know. Yet others may work in some other variant but not in the form in which they were implemented during the experiment. Finally, there are probably many interventions yet to be thought of that have the potential to change the world.²¹

Given this list of successful interventions, we ask, How much will it cost to more or less mechanically scale them up to a global level? Our definition of global covers only low-income countries (LICs) unless stated otherwise. We take the population of each country to be the average of the current population and the projected 2015 population.²²

We calculate costs by taking the point estimate of the per person cost for each program and adjust it for each country. This adjustment involves converting the expenditure on goods (vitamins, drugs) using the standard purchasing

power parity conversion factor, and scaling up (or down) the expenditures on services (teachers, health workers) in proportion to the gross domestic product per capita at current exchange rates.²³ For each country, the size of the targeted population is derived from demographic information about the country. The cost spreadsheet is available online.²⁴ Figures are normalized to year 2000 U.S. dollars.

The results are shown in table 2.4. Our calculations show that a recurring annual expense of about \$11 billion could be justified by the hard evidence we already have, without including PROGRESA. In addition, if we accept (nonexperimental) results in Cutler and Miller (2003) showing that the benefits from improved water supply are enormous, we should consider investing in water supply infrastructure. This will cost an additional (one-time) \$11 billion.

The \$11 billion number was our best guess in early 2004. Since then there have been a number of randomized evaluations that have reported success. These include a program in Kenya that gives incentives to girls to do more school and another program in Kenya that gives girls school uniforms as a way of reducing HIV infection rates (see www.povertyactionlab.org for details and other examples). As a result, if we had done this calculation today, the amount would be substantially larger.

Even \$11 billion is, however, a substantial amount. It is more than the World Bank gives out as IDA loans (the main form of World Bank aid) in a given year (an average of US\$6,154 million between 1996 and 2001). If we add to this what countries absolutely need to make essential macro adjustments of the kind discussed, there may be very little left from the total available domestic money.

To the extent that there is still money, it can be used to provide humanitarian aid. There are people in the world who are dying because they do not have enough money to buy food or medicines. Giving them money (or food or medicines) may not promote development, but it is hard to imagine that it would not be a good thing.

Indeed once we decide that we are willing to make cash gifts to people, we could make that gift conditional on the recipients' fulfilling certain conditions such as sending their children to school. This is what PROGRESA does.

It is true that this does not deal with most of the objections against giving away money. The targeting problem was solved in Mexico by using the bureaucracy to make sure that money goes to the right person, which is part of why it makes the program costly. But it is not clear that perfect targeting is worth the effort, given that most people in developing countries are actually quite poor. It may be better to set up rather lax criteria for eligibility so that only the rich

ineligible, and then to randomly check claimants and impose harsh punishments on those who are caught cheating. It is usually quite easy to identify the rich in poor countries (ownership of a car, for example, might be used). It is also true that we do not eliminate the possibility that the money may just go into consumption. But even this is less of a problem than it might appear. A number of studies, including those based on the PROGRESA experiment, have shown that more money in the hands of the female members of the family in poor countries does translate into better nutrition for children and better health care.

Ultimately, however, we should not need to argue that giving away money is without its problems. After all, it may still dominate trying to give away goods, which, as we have seen, is fraught with problems.

Clearly, being open to the idea of giving away money will make it much easier to find things to do. Just scaling up PROGRESA, by our calculations, will cost \$23 billion a year, which pushes up total annual expenditure on good programs to \$34 billion.

2.7 To Conclude: A New Challenge for the Millennium

We live in an age of aid pessimism. There is a strong, if rarely completely articulated, presumption that aid can at best help people survive, but it cannot promote development. The U.S. government's new initiative, the Millennium Challenge Account (MCA), is based on the idea that the whole idea of aid giving needs to be rethought. In particular, it wants to tie aid to country performance: only countries that pursue economic policies that the U.S. government approves of will be eligible for aid from this account. The premise is that aid has not been working because the policy environment is not right. While it is clear that this is a problem—there are countries where the risk of the money ending up in a government official's pocket is substantial—the thrust of our argument is that the way the money is planned to be spent is also a very big problem, but a problem whose source lies in the way the donor organizations function. Combined with the fact that many of the world's neediest live in the countries that will not make it onto the MCA list and that we expect the incentive effects of the MCA to be minimal, this suggests to us that the MCA approach amounts to abandoning a large part of the world's poorest for no fault of their own. A more effective and less unfair challenge may be to try to see if it is possible to design projects that work in the countries with the biggest problems. If we could make that work, we would not only help those who need it the most, but what is perhaps even more valuable, we will raise expectations and build hope where there is none.

Table 2.4
Program costs

Program (recurring annual expenditure)	Source and method of calculation	Randomized?	Cost (US\$2000 millions)
Education			
Remedial teaching based on the Batsakhi model developed by Pratham	Source: Banerjee et al. 2003. Calculation: $\text{Unit (cost per child-year), } BalkhashiCost$, from the Pratham 2001–2002 Reports (cited on June 28, 2003). Available at www.pratham.org/reports . We use the mean of costs reported for the Delhi, Mumbai, and Pune regions. Country cost = $BalkhashiPop * BalkhashiCost * GDPCorrection$	Yes	644
Universal education based on a 40:1 pupil-teacher ratio	Source: Angrist and Lavy 1999 ¹ Calculation: Country cost = $NotInSchool * SchoolCost$	No	1,544
School inputs (uniforms and textbooks)	Source: Kremer, Moulin, and Namunyu 2003. Calculation: Unit InputsCost from Kremer et al. (p. 44). Transportation costs ignored. Country cost = $InputsCost * SchoolAge * PPPCorrection$	Yes	2,268
Schooling vouchers	Source: Angrist et al. 2002. Calculation: Assumes that everyone is sufficiently motivated to achieve satisfactory performance, hence qualifying for the vouchers, and ignoring general equilibrium effects due to the resultant increase in private school fees. Unit $VoucherCost$ used in our calculations is the increase in public educational expenditure per lottery winner, given in Angrist et al. (1535). This is multiplied by four because at any time, there are four cohorts in high school. Country cost = $(4/15) * VoucherCost * ChildPop * GDPCorrection$	Yes	1,478
Monetary rewards to parents for sending children to school	Source: Behrman, Segupta, and Todd 2001. Calculation: Assumes that if the subsidy is large enough, everyone will want to send their children to school, and therefore everyone will get the subsidy. Unit $SubsidyCost$ is calculated from data given in Behrman et al. (2001, 1). This is multiplied by seven because at any time there are four cohorts getting the subsidy. Country cost = $(7/15) * SubsidyCost * ChildPop * GDPCorrection$	Yes	2,3142
Nutrition supplementation			
Iron	Source: Bobonis, Miguel, and Sharma 2004. Calculation: Unit $IronCost$ data from Miguel and Bobonis, private communication. The program covers five cohorts aged between ages two and six. Country cost = $(5/15) * IronCost * ChildPop * PPPCorrection$	Yes	346

Albendazole (deworming)	<p>Source: Kremer and Miguel 2004. Calculation: Unit DewormCost data from Miguel and Bobonis, private communication. The program covers five cohorts aged between ages two and six. Country cost = $(5/15) * \text{DewormCost} * \text{ChildPop} * \text{PPPCorrection}$</p>	Yes	36
Delivery of iron supplements and deworming pills through the Pratham Delhi Health Program model	<p>Source: See, e.g., Bobonis et al. 2003 Calculation: Unit TransportCost data from Miguel and Bobonis, private communication. The program covers five cohorts aged between ages two and six. Country cost = $(5/15) * \text{TransportCost} * \text{ChildPop} * \text{PPPCorrection}$</p>	Yes	270
Iodine	<p>Source: Cobra et al. 1997. Calculation: Unit IodineCost data from International Council for the Control of Iodine Deficiency Disorders. 1998. <i>IDD Newsletter</i> 14:3. The experiment involved a once-off supplement. Country cost = $(1/15) * \text{IodineCost} * \text{ChildPop} * \text{PPPCorrection}$</p>	Yes	12
Additional transportation costs	<p>Calculation: No handle on the transportation cost for the iodine intervention. Presumably some of it can be loaded on to the EPI program.</p>		
HIV			
Condom provision	<p>Source: Allen et al. 1992.</p>	Yes	135
Improving STD management	<p>Source: Grosskurth et al. 1995.</p>	Yes	428
Voluntary counseling and testing	<p>Source: Coates et al. 2000.</p>	Yes	116
Prophylaxis for opportunistic infections	<p>Source: Mwinga et al. 1998.</p>	Yes	40
Short-course zidovudine regime	<p>Source: Shaffer et al. 1999. Calculation: For all the HIV interventions: Main data source: Kumaranayake and Watts 2000. Her numbers are for sub-Saharan Africa (SSA), a group different from the low-income countries (LIC). The Commission of Macroeconomics and Health, Working Group 5 (2001, Paper 19) shows estimates of scaling up for both SSA and LIC for their set of interventions. Total Cost = $\text{Kumarayanayake SSA Cost} * \text{CMH LIC Cost} / \text{CMH SSA Cost}$</p>	Yes	4
Malaria: spraying	<p>Source: Rowland et al. 2000. Calculation: Malaria assumed to be prevalent in all the LICs. Unit SprayCost data from Vené et al. 1999. Country cost = $\text{Population} * \text{SprayCost} * \text{PPPCorrection}$</p>	Yes	1,627

Table 2.4
(continued)

Program (recurring annual expenditure)	Source and method of calculation	Randomized?	Cost (US\$2000 millions)
Fertilizer	Source: Dufo and Kremer 2003. Calculation: Agriculture statistics and fertilizer usage data from the 2003 World Development Indicators. We find the lowest <i>Fertilizer Consumption</i> (<i>100 grams per hectare of arable land</i>) and <i>Cereal Yield</i> (<i>kg per hectare</i>) between the four benchmark regions: [European Monetary Union], [High Income (nonOECD)], [High Income (OECD)] and the [United States]. We assume that a country is fertilizer deficient if both fertilizer consumption and the cereal yields are lower than the minimum value than the corresponding values (1045.875 and 2161.885) for the two benchmarks. For such countries we set <i>Deficient?</i> = 1, 0 otherwise. Unit <i>FertilizerCost</i> data from IFDC 2001. It does not have country-specific price estimates; hence, we are unable to correct for PPP differences. The recommended fertilizer consumption figure is also in the ballpark of the average recommendations of the net fertilizer (sum of different chemical compounds) across different countries, regions, and crops, as seen in Wichmann (n.d.) $\text{Country cost} = \text{Deficient?} * (1045.875 - \text{Fertilizer Consumption}) * \text{Land Area} * \text{Land Area Arable} (\% \text{ of total})$	Yes	1,848
Vaccination	Source: Hoke et al. 1988; O'Brien et al. 2003; Pérez-Schael et al. 1997. Calculation: Vaccine Fund 2002. It states that \$5,069 million of additional spending is required between 2001 and 2011. We divide this number by 11.	Yes	461
Total (excluding urban water provision and PROGRESA-style subsidies for school attendance)			11,257 million \approx 11 billion
Total (excluding urban water provision)			34,399 \approx 34 billion
Urban water provision	Source: Cutler and Miller 2003. Calculation: For more details, see the Water Construction section.	No	73
Total (including urban water provision)			34,472 million \approx 34 billion
One-off Infrastructure expenditure		R	One-time cost (US\$2000 million)
Urban water construction.	Source: Cutler and Miller 2003.	No	4,454

Rural water construction

No

10,867
 Calculation: For each continent, the World Health Organization and United Nation's Children Fund, 2000. *Global Water and Sanitation Assessment, 2000 Report*, New York: United Nation's Children Fund, provides the unit costs (*WaterCost*) of possible interventions for rural water construction (one off), urban water construction (one off), and urban water provision (per annum). Because of South Asia's huge size, we make a special effort to find cost information for this subcontinent (details on Web site). For our estimates, we choose to use the cheapest cost option. The assessment also provides country-level statistics on the Urban (*UrbanPop*) and Rural (*RuralPop*) population, and the percentage of urban (*U/NoAccess*) and rural (*R/NoAccess*) people without access to water.
 Urban country cost = $\text{PopCorrect} * \text{UrbanPop} * \text{U/NoAccess} * \text{UnitCost}$
 Rural country cost = $\text{PopCorrect} * \text{RuralPop} * \text{R/NoAccess} * \text{UnitCost}$

Total water construction

15,321 million ≈ 15 billion

Notes:

R: Randomized

1. Angrist and Lavy (1999) does not have a randomized experiment.

Assumptions behind the Table Numbers

Many bold and dubious assumptions have gone into the construction of this estimate of the cost of interventions for low-income countries. The fact that there may be economies of scale in expanding existing programs is ignored, as is the fact that the places that currently have these programs may not be randomly selected. Simplistic GDP and PPP corrections are used to correct for intercountry cost differences, and general equilibrium effects are ignored. We further assume linear population growth rates between 2000 and 2015 and that population is distributed equally among all ages between birth and fifteen years. While the time scale used here is the average annual cost between 2000 and 2015, future inflation is ignored. We try to err on the side of finding a low number.

Construction of Variables: The original data are from the 2002 Human Development Indicators in United Nations Development Program, 2002. *2002 Human Development Report*. New York: Oxford University Press (denoted in underlined italics, for example, *Total Population, 2000*), unless specified otherwise).

Country population estimate: $\text{Population} = 1,000,000 * (\text{Total population (millions)}, 2000 + \text{Total population (millions)}, 2015) / 2$
 Population growth factor estimate: $\text{PopCorrect} = (\text{Total population (millions)}, 2000 + \text{Total population (millions)}, 2015) / (2 * \text{Total population (millions)}, 2000)$

Child population estimate: $\text{ChildPop} = (1/100) * \text{Population under age 15 (as \% of total)}, 2000 * \text{Population}$
 Percentage who need remedial education (method 1) (assumes that completion rates are proportional to the number of students in school. In other words, with universal education, the remedial rate will drop): $\text{Remedial1} = 1 - \text{Primary completion rate} / \text{combined enrollment rate}$
 Percentage that needs remedial education (method 2) (assumes that remedial rate will stay constant even with full enrollment; children do not attend school because they are unable to cope): $\text{Remedial2} = (100 - \text{Primary completion rate}) / 100$
 Population who need the Balkashi program (a primary school length of six years is assumed): $\text{BalkashiPop} = (6/15) * \text{ChildPop} * (\text{Remedial1} + \text{Remedial2}) / 2$

Table 2.4
(continued)

Number of school-aged population not in school: $NotInSchool = (6/15) * ChildPop * (100 - CombinedEnrollmentData)/100$
 Cost per child year of education (wage assumed at 3.6 times GDP per capita, at a 40:1 teacher-pupil ratio, and allowing recurrent nonteacher costs to comprise 33 percent of the expenditure): $SchoolCost = 3.6 * 1.5 * GDP(US\$billions), 2000 / (TotalPopulation(millions), 2000 * 40)$
 Correction for costs of different services (we expect service costs to differ across countries and assume these costs are proportional to GDP per capita. Let the country that we use for the point estimate be country X). Let the country that we want to estimate costs for be country Y. Then: $GDPperCap = \frac{GDP(US\$billions), 2000}{TotalPopulation(millions), 2000}$
 $GDPCorrection = GDPperCap * [GDPperCap_X]$
 Correction for costs of different goods (we expect goods costs to differ across countries, and assume these costs are proportional to the country's PPP. Let the country that we use for the point estimate be country X). Let the country that we want to estimate costs for be country Y. Then: $PriceLevelUS = \frac{GDP(US\$billions), 2000}{GDP(PPPUS\$billions), 2000}$
 $PPPCorrection = PriceLevelUS * [PriceLevelUS_X]$
 Missing countries: We do not have detailed data for seven low-income countries. Hence, to extrapolate our cost estimates to include these countries, we make the following correction: Worldwide LIC costs = [Sum of country costs calculated (for nonmissing LIC countries)] * [Sum of LIC population (2002 estimate)] / [Sum of nonmissing LIC population (2002 estimate)]
 This correction ends up increasing our estimates by 3 percent.

Appendixes

Table 2.A.1
Evaluation organizations

Organization (evaluation office Web site)	Evaluation office	Year	Percentage of projects independently evaluated	Level of evaluation detail available online for independent evaluations	Rating scale for independent evaluations	Evaluation categories for independent evaluations
National Organizations						
Department for International Development (DFID) http://62.189.42.51/ DFIDstage/ policiesandpriorities/ files/ev_home.htm	Evaluation Department (EVD)	1968 ¹ onward	Very few: An aver- age of 9 projects were independently evaluated annually 1993–1999. ² Only 25% of bilateral expenditure is cov- ered by a project completion report (PCR) done by program managers. ³ “There is no system of independent veri- fication, although one has been mooted,” and “more significantly, PCRs are not used.” ⁴	All evaluation reports and their summaries are online or can be ordered online.	Numerical rating scales for numerous dimensions for the PCRs, but these are not independent evaluations. Independent obser- vations are mostly prose based. Since independent evalua- tions are carried out by different evalua- tors, no standardized rating scale/category exists.	

Table 2.A.1
(continued)

Organization (evaluation office Web site)	Evaluation office	Year	Percentage of projects independently evaluated	Level of evaluation detail available online for independent evaluations	Rating scale for independent evaluations	Evaluation categories for independent evaluations
U.S. Agency for International Development (USAID) http://www.dec.org/partners/eval.cfm	Bureau for Policy and Program Coordination, ⁵ Center for Development Information and Evaluation (CDIE)	Before 1995 ⁶ After 1995 to 2000 After 2000 ⁹	All projects were evaluated, ⁷ but how many are independent? Regarded poorly. ⁸	Individual project evaluations. Mostly joint authored with country mission, raising questions about independence. Annual report uses country macro- economic values as performance indicators. ¹¹	Qualitative measures used in its performance and accountability report. ¹²	

Regional Development Banks					
African Development Bank (AfDB) http://www.afdb.org/about_adb/OPEV.htm	Operations Evaluation Office	1964 (est.)–2001 2001 onward ¹⁴	Annual reviews of development effectiveness starting in 2003. ¹⁵ Annual review of results of operations evaluation apparently published, according to http://www.afdb.org/about_adb/OPEV_evaluation_guidelines.htm , however, not available online.	Project evaluations prior to 1997 are available at the DAC Web site. ¹⁶ Qualitative abstracts of project performance audit reports from 1994 to 1997 available online or by request to the webmaster. However, it is not clear if these were independent evaluations.	ARDE 2003 not currently available online.
Asian Development Bank (ADB) http://www.adb.org/Evaluation/	Operations Evaluation Department	1974 onward ¹⁷	40% evaluated ¹⁸	Individual project reports (from 1995 onward). ¹⁹ Quantitative panel data grouped by sector and region available. ²⁰ Summary tables collating evaluation details per project evaluated. ²¹	Overall project rating: Highly Successful, Partly Successful, Unsuccessful ²²
					Relevance, Effectiveness, Efficiency, Sustainability, Institutional Development (REES) ²³

Table 2.A.1
(continued)

Organization (evaluation office Web site)	Evaluation office	Year	Percentage of projects independently evaluated	Level of evaluation detail available online for independent evaluations	Rating scale for independent evaluations	Evaluation categories for independent evaluations
European Bank for Reconstruction and Development (EBRD) http://www.ebrd.org/projects/eval/index.htm	Project Evaluation Department	1991 (year of establishment) ²⁴ to 2003	44% (1991 to 1998) ²⁵	Examples of successful and less successful projects. Time series quantitative data of performance for all projects grouped together. ²⁶ Forty project summaries and lessons learned (each one page long) representing "a cross-section of EBRD investment operations" published online. Previous year overview reports and all of the other project completion reports are internal documents. ²⁷	Quantitative rating scales (> 3 possible ratings) for each category ²⁹	Transition impact, Environmental performance of sponsor and bank, extent of environmental change
		2003 onward	76% (1993 to 2002) ²⁸	In 2003, for the first time, Annual Evaluation Overview Report available online, giving quantitative evaluation		

<p>Inter-American Development Bank (IADB) http://www.iadb.org/cont/evo/evo_eng.htm</p>	<p>Evolve over time. (1) Group of Three Controllers, (2) Office of External Review and Evaluation and the Operations Evaluation Office, (3) Office of Evaluation³¹</p>	<p>1959-1998</p>	<p>Seldom undertaken, even then, only by borrower.³² Exact figures unknown, but we find that no ex post evaluation was carried out for Mexico throughout the 1990s,³³ which gives cause for worry.</p>	<p>Six to seven country programs evaluated each year (at least from 1999 to 2001). However, the lack of ex post program evaluation in the 1990s means that countries are evaluated on macro-performance. The presence of multiple donors makes assigning credit hard.³⁶ Very few project evaluations are done independently.³⁷</p>	<p>Country program evaluations, each covering a ten-year period.³⁸ No cross-country or cross-sector comparisons.</p>	<p>Quantitative figures, but these are based on project evaluations that were primarily self-evaluated,³⁹ and sometimes based on the most recent project performance monitoring reports (as opposed to the project completion reports, which were seldom done⁴⁰)</p>	<p>Relevance, coherence, efficiency and effectiveness⁴¹</p>
<p>Office of Evaluation and Oversight³⁴ (OVE)</p>	<p>1999 onward³⁵</p>						

Table 2.A.1
(continued)

Organization (evaluation office Web site)	Evaluation office	Year	Percentage of projects independently evaluated	Level of evaluation detail available online for independent evaluations	Rating scale for independent evaluations	Evaluation categories for independent evaluations
International Organizations						
International Monetary Fund (IMF) http://www.imf.org/external/np/ieo/index.htm	Office of Internal Audit and Inspection Evaluation Group of Executive Directors ⁴²	Before 1996 Between 1996 and 2000 After 2000	Insignificant numbers of external evaluation Trial run that covered a range of topics and were of different scale. ⁴³ At capacity, five projects to be undertaken each year, including both country specific cases ⁴⁴ and broader thematic questions. ⁴⁵	Commitment to promptly publish all reports, unless under exceptional circum- stances. ⁴⁶ No cross- sector (for example, comparing the per- formance of capital account crisis inter- vention versus the pro-longed use of IMF resources) or time series compari- sons available yet.	While not explicit, key cross-country macroeconomic indicators are provided and implicitly used as indicators of per- formance. Within the two reports published thus far, there is an attempt to dis- tinguish between different levels of success.	As of January 2004, four reports were published. ⁴⁷ These looked at surveillance, pro- gram effective- ness, and IMF governance. The medium program is given in its annual report. ⁴⁸

<p>United Nations Development Program (UNDP) http://www.undp.org/ee/</p>	<p>Office of Evaluation and Strategic Planning (OESP)⁴⁹ Evaluation Office (EO)</p>	<p>Before 1996 1996-1999</p>	<p>For programs less than US\$1 million, country managers decide if project should be evaluated. Mandatory evaluations for projects over US\$1 million.⁵⁰ Even then, the compliance for mandatory evaluations was less than 80%.⁵¹</p> <p>Requirement that all projects over US\$1 million be evaluated is abolished.⁵²</p> <p>Decentralized evaluation process, with country managers selecting the evaluation teams and designs the term of reference.⁵³ The EO concentrates on independent, country-level assessment of development results (five to ten countries a year), and sector-thematic level evaluations.⁵⁴</p>	<p>Individual project reports available online.⁵⁵ Quantitative time series data for each performance indicator⁵⁶ (all projects lumped together), and cross-sector (for projects evaluated from 1999 to 2000) performance impact indicators.⁵⁷</p>	<p>Yes/partial/no and Significant/Satisfactory/Poor⁵⁸</p>	<p>Typically covers relevance, performance and success (impact, sustainability, and contribution to capacity building).⁵⁹ However, this criterion is different from the previous year, which raises some questions.⁶⁰</p>
--	--	---------------------------------------	--	--	--	---

Table 2.A.1
(continued)

Organization (evaluation office Web site)	Evaluation office	Year	Percentage of projects independently evaluated	Level of evaluation detail available online for independent evaluations	Rating scale for independent evaluations	Evaluation categories for independent evaluations
World Bank, http://www.worldbank.org/ocd/	Operations Evaluation Department	1973 ⁶¹ onward	Independent evaluation at the country, sector and project levels. Project performance assessment reports for 25% of all completed projects. ⁶²	Individual project reports. ⁶³ Quantitative panel data grouped by sector and region available. ⁶⁴ Summary tables collating evaluation details per project evaluated. ⁶⁵	Outcome: Highly Satisfactory (Sat), Sat, Moderately Sat, Moderately UnSat, UnSat, Highly UnSat, Sustainability: Highly Likely, Highly Likely, Highly Unlikely, Unlikely, Unlikely Institutional Development: Substantial, Modest, Negligible ⁶⁶	Outcome, Sustainability, Institutional Development ⁶⁷

Note: We look at only the level of evaluation detail available online for the latest incarnation of the evaluation office.

Notes for Appendix 2.A.1

1. Year of first independent evaluation as given by Department of International Development. 2003. *Catalogue of DFID Evaluation Studies*. DFID Evaluations Department. Available at http://www.dfid.gov.uk/Pubs/files/eval_studies_catalogue.pdf.
2. Flint, Michael, et al. 2002. "How Effective Is DFID? Development Effectiveness Report 2001," 2nd Draft. DFID, March 30, paragraph 118.
3. *Ibid.*, paragraph 115.
4. *Ibid.*, paragraph 116.
5. United States Agency for International Development. "Agency Reorganization: Generic Functional Statements." Cited 11 Feb 2004. Available <http://www.usaid.gov/about/reform/functions.html>.
6. Clapp-Wincek, Cynthia, and Richard Blue. 2001. "Evaluation of Recent USAID Evaluation Experience." Working Paper 320. Washington: Center for Development Information and Evaluation

8. *Ibid.*, p. 37. Here we face the related problems of few evaluations done ("Of most concern is the very limited number of in-depth, program evaluations"), lack of learning even if evaluations are carried out ("Most of the evaluation work that is being done is being done by partners. The partner organizations are learning from the experience; USAID is not"), lack of independence (the USAID managers decide if they want to do the evaluations), the fly-in approach ("Scopes ask a team to come for 4-6 weeks and interview the mission, the activity staff, and 'representatives' of the local people. There isn't enough time to get any kind of representative sample. The team frequently tells the USAID manager pretty much what he already knows.")
9. Quote from *ibid.*, p. 1, "The ADS 200 series (Sept. 2000) added a new dimension to evaluations. The 'Reform Vision' in ADS 200 states these expectations: 'Applying the lessons of successes and failure systematically and providing leadership in tackling complex problems that demand multi-agency or multi-donor responses.'" 10. DuRette, Jean, and Glenn Slocum. 2001. *The Role of Transitional Assistance: The Case of East Timor*. Washington, D.C.: U.S. Agency for International Development.
11. Center for Development Information and Evaluation. 2001. *FY 2000 Performance Overview*. Washington, D.C.: U.S. Agency for International Development. See p. 29, for example.
12. U.S. Agency for International Development. 2002. *USAID Performance and Monitoring Report, FY 2002*. Washington, D.C.: United States Agency for International Development. See appendix 2.
13. African Development Bank Group. "The ADB Group—In Brief." Cited February 11, 2004. Available at http://www.afdb.org/knowledge/documents/ADB_in_brief.htm.
14. U.S. Department of the Treasury. 2003. "African Development Bank Group." Cited 11 Feb 2004. Available at http://www.ustreas.gov/offices/international-affairs/intl/fy2003/tab10_afdbg_afdb.pdf.
15. African Development Bank Group. 2002. "Strategic Plan 2003-2007." Available at http://www.afdb.org/knowledge/publications/pdf/adb_strategic_plan2003-2007e.pdf. See p. 52.
16. See <http://www.dac-evaluations-cad.org/dac/>.
17. Asian Development Bank. Operations Evaluation Department. 2003. *Annual Review of Evaluation Activities in 2002*. Manila: Asian Development Bank. See p. 61.
18. Asian Development Bank. "Frequently Asked Questions." Cited February 11, 2004. Available at <http://www.adb.org/Evaluation/faqs.asp>.
19. All reports online since 1995, at Asian Development Bank: <http://www.adb.org/Evaluation/reports.asp>.
20. Asian Development Bank. Operations Evaluation Department. 2003. *Annual Review of Evaluation Activities in 2002*. Manila: Asian Development Bank. See p. 67.
21. *Ibid.*, p. 47.
22. *Ibid.*
23. Asian Development Bank. "PPMS Project Framework and Performance Indicators 2." Slide show. Available at http://www.adb.org/Documents/Slideshows/PPMS/4b_PPMS_Indicators.pdf.
24. Project Evaluation Department. 1999. "Project Evaluation Department." London: European Bank for Reconstruction and Development. See p. 6.
25. *Ibid.* See p. 6.
26. European Bank for Reconstruction and Development. 2003. *Annual Report 2002: Annual Review and Financial Report*. London: European Bank for Reconstruction and Development. See p. 70.
27. Fredrik Korfker, corporate director, evaluation, EBRD, September 9, 2003. private communication. 28. European Bank for Reconstruction and Development. Project Evaluation Department. 2003. *Annual Evaluation Overview Report 2003*. London: European Bank for Reconstruction and Development. See p. 1.

49. United Nations Development Program and the United Nations Population Fund. Executive Board. 2000. "Evaluation: Report of the Administrator." Available at <http://www.undp.org/excecbrd/pdf/dp00-34e.pdf>. See p. 2.
50. United Nations Development Program, Evaluation Office. 2001. "Development Effectiveness: Review of Evaluative Evidence." New York: United Nations Development Program. See p. 15.
51. United Nations Development Program and the United Nations Population Fund, Executive Board. 2001. "Evaluation: Report of the Administrator." Available at <http://www.undp.org/excecbrd/pdf/dp01-26e.pdf>. See table 6.
52. United Nations Development Program, Evaluation Office. 2002. "Handbook on Monitoring and Evaluating for Results." New York: United Nations Development Program. See part 1, table 3.
53. *Ibid.* See part 2, p. 45.
54. United Nations Development Program, Evaluation Office. 2001. See p. 15.
55. See United Nations Development Program Evaluations Database. <http://www.undp.org/eo/database/index.htm>.
56. United Nations Development Program, Evaluation Office. 2001. See pp. 15-19.
57. *Ibid.* See p. 35.
58. *Ibid.* See p. 19.
59. *Ibid.* See p. 15.
60. *Ibid.* See p. 10.
61. World Bank, Operations Evaluation Department. 2002. *OED: The First 30 Years*. Conference Proceedings, September 23, 2002. Available at http://www.worldbank.org/wbi/B-SPAN/docs/oed_thirty.pdf.
62. World Bank, Operations Evaluation Department. 2002. *2001 Annual Review of Development Effectiveness*. Washington, D.C.: World Bank. See annex D.
63. See <http://hweb18.worldbank.org/servlet/OEDSearchServlet?SearchType=byField&PerPage=20&DbURL=oed/oeedoclib.nst&Series=OED%20Project%20Evaluation>.
64. World Bank, Operations Evaluation Department. 2003. *2002 Annual Review of Development Effectiveness*. Washington, D.C.: World Bank. See table 5.
65. *Ibid.* See table 14.
66. *Ibid.* See table 14.
67. *Ibid.* See table 14.

Table 2.A.2

Table of Interventions

The following abbreviations are used in this table:

C: Intervention Category

E: Education

P: Direct poverty reduction

H: Health

O: Others

S: Is the program self-sustainable? We claim that a proposed program is not self-sustainable if the researchers ensured compliance amongst the test subjects. This is an issue for the clinical evaluations of nutritional supplements which look at the impact of a supplement assuming compliance. In real life, compliance cannot be assumed.

S: Randomized?

Intervention	C	Evaluation cited	S	R	Benefits	Including cost?
Water and Sanitation						
Clean water through a water container with a cover and a sprout	H	Roberts et al. 2001 Self-sustainable because Malawian field worker used, and "this [how to use the bucket] educational message generally took less than one minute and was never reinforced or restated during the study." Not scaled up because it was carried out in a specific instance of a refugee camp that had experienced a cholera outbreak.	Yes	Yes	Reduced diarrhea	No
Latrine provision	H	Daniels et al. 1990 Selection bias: Ownership depends on constituency, and distance to the recruiting health facility.	Yes	No	Reduced diarrhea	No. Not random
Latrine provision	H	Estey 1996	Yes	No	Reduced diarrhea, taller and heavier children	No. Not random
Education for water sanitation behavior	E H	Potential nonrandom location of latrines. Stanton 1987	Yes	Yes	Reduced diarrhea	Yes. (We assume that if everyone goes to school, the schools will do the education.)
Historical clean water interventions in America	H	Cutler and Miller 2003 Attempts to deal with potential endogenous placement.	Yes	No	Clean water reduces mortality	Yes (although not random)

Health care services		Summary: Gelband and Stansfield 2001			
		Yes	No	Better skills	No. Not random
Midwife services	H	Walker et al. 2002 Not random due to participation selection bias.			
Midwife services	H	Frankenberg and Thomas 2001		Increased women Body Mass Index	No. Not random
Reduction in antenatal care	H	Actual program, but selective nonrandom midwife placement. Munjanja, Lindmark, and Nystrom 1996		No effect on maternal/fetal outcomes	Cost reduction, not included
Reduction in antenatal care	H	Actual reduction in number of visits, clearly sustainable. Villar et al. 2001		No effect on maternal/fetal outcomes	Cost reduction, not included
Home-based neonatal care	H	Actual reduction in number of visits, clearly sustainable. Bang et al. 1999		Avert 1 death per 18 neonates cared for	No
		Control villages were not randomly chosen: Villages were used as controls because suitable women could not be found, or population was fewer than 300.			
		Summary: Behrman and Knowles 2003; Nemer, Gelband, and Jha 2001			
Nutritional supplements					
Deworming drugs	E	Kremer and Miguel 2004		Improved health and school participation, even in neighborhood schools	Yes
	H	Self-sustainable because "medical treatment was delivered to schools by Kenya Ministry of Health public health nurses and ICS public health officers" and this was done within the community setting (we see some children not participating).			
Iron supplementation	E	Bobonis, Miguel, and Sharma 2003		Increased participation, reduced absenteeism	Yes
	H	Existing preschool network used. Brown et al. 2002		Positive responses in health and weight	No. Not program
Zinc	H	Not self-sustainable because of overtly heavy fieldworker involvement—"in most cases, confirmation that the supplements were successfully delivered to the study subjects."			

Table 2.A.2
(continued)

Intervention	C	Evaluation cited	S	R	Benefits	Including cost?
Oral iodized oil supplementation	H	Claudine et al. 1997	Yes	Yes	Improved infant survival	Yes
Vitamin A supplementation	H	Grotto et al. 2003	No	Yes	"No consistent overall protective effect on the increase of diarrhea"	No
Supplements for lactating women	H	Many clinical trials carried out over the years. The above meta-analysis shows that the net impact is not consistently positive. Tinker, Finn, and Epp 2000 Supplements were provided daily on a <i>volunteer</i> basis at a centralized location—there was no compulsion to consume it.	Yes	No	Reduced low-weight babies	No, Not random
Folate supplementation (literature review)	H	Mahomed 1997	—	—	Not enough evidence to evaluate effectiveness on clinical outcomes	No
Antiplatelet supplementation (literature review)	H	Knight et al. 2000	—	—	Timing and dosage knowledge insufficient	No
Supplemental feeding (literature review)	H	Rush 2000	—	—	Insufficient knowledge to decide if nutritional supplements are good overall.	No
Tuberculosis		Summary: Borgdorff, Floyd, and Brockmans 2001.				
Short-course chemotherapy	H	China Tuberculosis Control Collaboration. 1996.	Yes	No	Increased cure rates	No, Not random
BCG	H	Tuberculosis Research Center (ICMR), Chennai 1999.	Yes	Yes	No effect of BCG on TB	No, No effect
Preventive therapy	H	Mwinga et al. 1998	Yes	Yes	Drugs reduced TB infection for HIV-infected people in Zambia	Yes (under HIV interventions)

Vaccinations	H	Here we look for evidence for the new vaccines that the Global Alliance for Vaccines and Immunization is trying to introduce. The Vaccine Fund 2002, <i>Strategic Plan 2002-2006</i> . There is a strong evidence base for the more established basic vaccines. Here we assume that all the vaccines can be implemented as programs, due to the success of the Expanded Program for Immunization.	Yes	Yes	Reduced encephalitis attack rate	Yes
Japanese encephalitis vaccine	H	Hoke et al. 1988	Yes	Yes	Prevents vaccine serotype invasive pneumococcal disease	Yes
Pneumococcal vaccine	H	O'Brien et al. 2003	Yes	Yes	Protection against severe diarrhea and dehydration, and reduced hospital admissions	Yes
Quadrivalent vaccine	H	Pérez-Schael et al. 1997	Yes	Yes		
Malaria						
Indoor spraying	H	Summarized in Meek, Hill, and Webster 2001 Rowland et al. 2000	Yes	Yes	Reduction in anopheline porous rates	Yes
Insecticide-treated nets	H	Shulman et al. 1998	Yes	Yes	No significant impact	No. Not significant
Outdoor spraying	H	Cutler 2003	Yes	No	Increase in arable land	No. Not random
HIV						
Voluntary HIV-1 counseling and testing	H	Summarized in Jha et al. 2001; Behrman and Knowles 2003 Coates et al. 2000	Yes	Yes	Reduced intercourse with nonprimary partners	Yes
Condom provision in motel rooms	H	Egger et al. 2000	Yes	Yes	Increased condom use	Yes
Package, including drug supply, health education and STD reference clinic	H	Grosskurth et al. 1995	Yes	Yes	Reduced HIV incidence	Yes
Home-based mass antibiotic treatment	H	Wawer et al. 1999	Yes	Yes	No effect	No

Table 2.A.2
(continued)

Intervention	C	Evaluation cited	S	R	Benefits	Including cost?
Confidential HIV testing and condom promotion	H	Allen et al. 1992	Yes	Yes	Increased condom use, reduced rates of gonorrhoea and HIV in urban Rwandan women	Yes
Short-course zidovudine for babies who are not breast-fed	H	Shaffer et al. 1999	Yes	Yes	Reduced risk of mother-to-child HIV transmission	Yes
Short-course zidovudine for breast-fed babies	H	Wiktor et al. 1999	Yes	Yes	Reduced risk of mother-to-child HIV transmission	Yes
Monetary transfers						
		Unlike the health-based interventions, all monetary transfers actually happened in a community-based context; hence, they must be self-sustainable.				
Decentralized targeting	P	Galasso and Ravallion 2001	Yes	No	Results better in more favorable conditions	No. Not program evaluation per se
Old age pension	H	Duflo 2003	Yes	No	Increased weight of girls	No. Not random
Poor-area development programs	P	Jalan and Ravallion 1998	Yes	No	Enough to prevent decline, but not enough for convergence	No. Not random
Microfinance						
		Unlike the health-based interventions, all micro-finance interventions actually happened in a community-based context; hence, they must be self-sustainable.				
Village-level microfinance	P	Kaboski and Townsend 2002	Yes	No	Mixed	No. Not random, not effective
Village-level microfinance	P	Morduch 1998	Yes	No	No	No. Not effective
School inputs						
		Unlike the health-based interventions, all school-based interventions actually happened in a community-based context; hence, they must be self-sustainable.				

School construction	E	Duflo 2004	Yes	No	Increase labor force participation, but reduce wages of older cohorts	No. Not random
Flip charts	E	Glewwe et al. 2000	Yes	No/Yes	No evidence with prospective, positive with retrospective	No. No evidence using random evaluation
Uniforms	E	Kremer, Moulin, and Namunyu 2003	Yes	Yes	Reduce dropout rates, without reducing test scores	Yes
Teacher incentives	E	Glewwe, Ilias, and Kremer 2003	Yes	Yes	No evidence, teachers teach to test	No. Not successful
Textbook	E	Glewwe, Kremer, and Moulin 2003	Yes	Yes	Raised test scores of the highest quintile, more likely to go to secondary school	No. Very skewed results
Remedial education program	E	Banerjee et al. 2003	Yes	Yes	Increased learning	Yes
Teacher provision	E	Chin 2002	Yes	No	Increased female primary school completion and literacy	No. Not random
School incentives						
School meals	EH	Vermeeersch 2002	Yes	Yes	Conditional test score improvement, higher participation, cut into instruction time	No. Questionable effectiveness
PROGRESA	E	Schultz 2001	Yes	Yes	Higher enrollments	Yes
PROGRESA	E	Behrman, Segupta, and Todd 2001	Yes	Yes	Increased educational attainment, % attending junior secondary school	
PROGRESA	H	Behrman and Hoddinott 2001	Yes	Yes	0/- effect with means, + stature with Fixed Effects	
PROGRESA	H	Gerler and Boyce 2003	Yes	Yes	Health improvements	

Table 2.A.2
(continued)

Intervention	C	Evaluation cited	S	R	Benefits	Including cost?
School vouchers	E	Angrist et al. 2002.	Yes	Yes	Higher completion, less repeating of grades	Yes
Hygiene education	EH	Haggerty et al. 1994	Yes	Yes	Reduced diarrhea	Yes (folded into other education estimates)
Fertilizer						
Fertilizer adoption through an NGO program		Duollo and Kremer 2003	Yes	Yes	Increased adoption, high rate of return	Yes
Roads						
Rural road construction		van de Walle 2004	Yes	No	Research in progress	Not random, research in progress

Notes

1. Agriculture, Electric Power and Other Energy, Environment, Mining, Oil and Gas, PSD/Industry, Telecommunications and Informatics, Transportation, Urban Development, and Water Supply and Sanitation.
2. Others, like USAID, just report their assessment of how well the country is doing.
3. The European Bank for Reconstruction and Development published its results online for the first time in late 2003. Even though it provides item-by-item ratings for all projects evaluated since 1993, it does not provide the information needed to match these projects to funding information, and therefore the data are not yet usable.
4. Percentage figures from the annual reports of the World Bank: Summaries of Projects Approved for IBRD, IDA, and Trust Funds in each Fiscal year. For each project, the following information is provided: country name, sector, brief project description, World Bank contribution, and total project cost.
5. Sectors were determined by the classifications in World Bank (2000b, Annex 1), which conveniently classifies the historical performance data into the 1990–1993 and 1994–1997 periods. From these sectors, the “Social” and “Environmental” sectors were dropped because zero/one evaluation was done between 1990 and 1993. We do not have 1987–1990 data for “Mining,” “Multi-sector” and “Public Sector Management” due to a change in sector classification.
6. The 1987–1990 performance data are derived by collating information from World Bank (1989, 1993). The years 1987 and 1990 here refer to the year that the project was evaluated; hence, they correspond to projects that had slightly earlier exit fiscal year groups. The 1990–1993 and 1994–1997 performance data are derived from World Bank (2000a), and correspond to projects that had exit fiscal years within the respective ranges. The 1998–2001 data are derived from collating information from World Bank (2000a, 2002a).
7. These are the four selected sectors in Morra and Thumm (1997, table 1.32).
8. The Technical Assistance category was dropped in World Bank (1994).
9. Morra and Thumm (1997, table 1.32).
10. Annual reports of the World Bank.
11. Here, unlike in the results in the previous table, we are using the gap in performance between t and $t - 1$ rather than that between $t - 1$ and $t - 2$, because of data limitations in the case of the World Bank. For the ADB we could use the gap between $t - 1$ and $t - 2$. The results are very similar.
12. For both the ADB and the World Bank, the sector dummies as a group are significant at the 1 percent level (based on their joint F -statistic), as are the country dummies, though the sector dummies become insignificant for the ADB when we drop the multisector/others category.
13. Data from World Bank (1998, 1999, 2000a, 2001, 2002a, 2003a).
14. Bill Easterly, another long-term bank employee, makes a similar point in an article in the *Journal of Economic Perspectives* (Easterly 2003).
15. Those who are interested in the argument for (and against) randomized trials as a basis for social policy may want to see Duflo (2004) and Duflo and Kremer (2004).
16. See Angrist and Lavy (1999) for an example of a very convincing natural experiment.
17. This is, in effect, how Stanley Fischer, a former chief economist of the bank, put it while commenting on our previous paper at the meetings of the American Economic Association.
18. We feel that fieldworkers ensuring that villagers comply with their daily supplement dosage constitutes overly strong intervention.
19. For example, this was the basis for choosing Miguel and Kremer’s (2004) work on deworming and Bobonis, Miguel, and Sharma (2003) on iron supplementation.
20. One place where we did exercise some judgment is in including the study by Angrist and Lavy (1999), which used a natural experiment in Israel to estimate the effect of class size on learning. This

natural experiment was based on the fact that in Israel, class size is capped at forty (the so-Maimonides rule), which generates a sharp discontinuity whenever class size hits forty. The study exploits the fact that this discontinuity generates something very close to a pure random assignment across schools.

21. We also recognize that researchers do not always have control over the experimental design, and, in any case, that the research might have been done with different objectives in mind (for example, researchers might be interested in knowing the impact of supplements on growth if it were assumed that compliance was not going to be a problem). Inclusion in the highlighted list does not reflect any opinion of the quality of the research or the researcher. It merely reflects the suitability of the research for our purposes.

22. We chose 2015 because it is when the Millennium Development Goals are meant to be achieved and many of our programs could be ways to achieve the goals.

23. Under the plausible assumption that the PPP correction is unreliable for the salaries of service providers, who tend to be scarce in developing countries.

24. <http://web.mit.edu/ruimin/www/whatworks/whatworksest.xls>

References

- Allen, S., et al. 1992. "Confidential HIV Testing and Condom Promotion in Africa. Impact on HIV and Gonorrhea Rates." *Journal of the American Medical Association* 268, no. 23 (December): 3338–3343.
- Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement." *Quarterly Journal of Economics* 114, no. 2 (February): 533–575.
- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Effects of Private Schooling in Columbia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92, no. 5 (December): 1535–1558.
- Asian Development Bank. 2003. *Annual Review of Evaluation Activities in 2002*. Available at www.adb.org/Documents/PERS/RPE_OTH_2003_12.pdf.
- Banerjee, Abhijit, and Ruimin He. 2003. "The World Bank of the Future." *American Economic Review* 93, no. 2 (May): 39–44.
- Banerjee, Abhijit, et al. 2003. "Remedying Education: Evidence from Two Randomized Experiments in India." Poverty Action Lab Paper 4.
- Bang, A. T., et al. 1999. "Effect of Home-Based Neonatal Care and Management of Sepsis on Neonatal Mortality: Field Trial in Rural India." *Lancet* 354, no. 9194 (December): 1955–1961.
- Behrman, Jere R., and James C. Knowles. 2003. "Assessing the Economic Returns to Investing in Youths in Developing Countries." Mimeo, University of Pennsylvania.
- Behrman, Jere, and John Hoddinott. 2001. "Program Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition." Penn Institute for Economic Research, Working Paper 02-006.
- Behrman, Jere, Pilali Segupta, and Petra Todd. 2001. "Progressing Through PROGRESA: A Impact Assessment of a School Subsidy Experiment." Penn Institute for Economic Research, Working Paper 01-033.
- Bobonis, Gustavo, Edward Miguel, and Charu Sharma. 2004. "Iron Deficiency Anemia and School Participation." Mimeo, University of California, Berkeley.
- Brown, K. H., et al. 2002. "Effect of Supplemental Zinc on the Growth and Serum Zinc Concentrations of Prepubertal Children: A Meta-analysis of Randomized Controlled Trials." *American Journal of Clinical Nutrition* 75, no. 6 (June): 1062–1071.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2001. "Women as Policy Makers: Evidence from an India Wide Randomized Experiment." Working paper 8615, December. Cambridge, Massachusetts: National Bureau of Economic Research.

- Chin, Aimee. 2002. "The Returns to School Quality When School Quality Is Very Low: Evidence from Operation Blackboard in India." Mimeo, University of Houston.
- China Tuberculosis Control Collaboration. 1996. "Results of directly observed short-course chemotherapy in 112,842 Chinese patients with smear-positive tuberculosis." *Lancet* 347, no. 8998 (February): 358–362.
- Coates, et al. 2000. "Efficacy of Voluntary HIV-1 Counseling and Testing in Individuals and Couples in Kenya, Tanzania, and Trinidad: A Randomized Trial." *Lancet* 356, no. 9224 (July): 103–112.
- Cobra, Claudine, et al. 1997. "Infant Survival Is Improved by Oral Iodine Supplementation." *Journal of Nutrition* 127, no. 4 (April): 574–578.
- Commission of Macroeconomics and Health, Working Group 5. 2001. *Various Working Papers*. Geneva: World Health Organization. Available at http://www.cmhealth.org/cmh_papers&reports.htm.
- Cutler, David, and Grant Miller. 2003. "Clean Water Measures in American History." Private communication.
- Cutler, David. 2003. "The Economic Impacts of Public Health Improvements." Personal communication.
- Daniels, D. L., et al. 1990. "A Case-Control Study of the Impact of Improved Sanitation on Diarrhoea Morbidity in Lesotho." *Bulletin of the World Health Organization* 68, no. 4:455–463.
- De Cock, Kevin, Mary Glenn Fowler, Eric Mercier, Isabelle de Vincenzi, et al. 2000. "Prevention of Mother-to-Child HIV Transmission in Resource-Poor Countries: Translating Research into Policy and Practice." *Journal of the American Medical Association* 283, no. 9 (March): 1175–1182.
- Duflo, Esther, and Michael Kremer. 2003. "Understanding Technological Choices: Fertilizers in Western Kenya." Mimeo.
- Duflo, Esther, and Michael Kremer. 2005. "Use of Randomization in the Evaluation of Development Effectiveness." In *Evaluating Development Effectiveness*, ed. George Keith Pitman, Osvaldo N. Feinstein, and Gregory K. Ingram. New Brunswick, N.J.: Transaction Publishers.
- Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91, no. 4 (September): 795–813.
- Duflo, Esther. 2001. "The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia." Forthcoming in *Journal of Development Economics*.
- Duflo, Esther. 2003. "Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa." *World Bank Economic Review* 17, no. 1 (June): 1–25.
- Duflo, Esther. 2004. "Scaling Up and Evaluation." In *Proceedings of the Annual World Bank Conference in Development Economics Conference*. Washington, D.C.: World Bank, 341–369.
- Duflo, Esther, Pascaline Dupas, Michael Kremer, and Samuel Sinei. "Education and HIV/AIDS Prevention: Evidence from a Randomized Evaluation in Western Kenya." Mimeo. MIT.
- Easterly, William. 2001. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge, Mass.: MIT Press.
- Easterly, William. 2003. "Can Foreign Aid Buy Growth." *Journal of Economic Perspectives* 17, no. 3 (Summer): 23–48.
- Egger, M., et al. 2000. "Promotion of Condom Use in a High-Risk Setting in Nicaragua: A Randomized Controlled Trial." *Lancet* 355, no. 9221 (June): 2101–2105.
- Esrey, S. A. 1996. "Waste, Waste, and Well-Being: A Multicountry Study." *American Journal of Epidemiology* 43, no. 6 (March): 608–623.
- Frankenberg, Elizabeth, and Duncan Thomas. 2001. "Women's Health and Pregnancy Outcomes: Do Services Make a Difference?" *Demography* 38, no. 2 (May): 253–265.
- Galasso, Emanuela, and Martin Ravallion. 2001. "Decentralized Targeting of an Anti-Poverty Program." Forthcoming, *Journal of Public Economics*.

- Gertler, Paul, and Simone Boyce. 2003. "An Experiment in Incentive-Based Welfare: The of PROGRESA on Health in Mexico." Royal Economic Society Annual Conference 2003.
- Glewwe, Paul, et al. 2000. "Retrospective vs. Prospective Analyses of School Inputs: The Flip Charts in Kenya." NBER Working Paper 8018.
- Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2003. "Textbooks and Test scores: E from a Prospective Evaluation in Kenya." Mimeo, Harvard University.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2003. "Teacher Incentives." NBER W Paper 9671.
- Grosskurth, H., et al. 1995. "Impact of Improved Treatment of Sexually Transmitted Disease HIV Infection in Rural Tanzania: Randomized Controlled Trial." *Lancet* 346, no. 8974 (A 530–536).
- Grotto, I., et al. 2003. "Vitamin A Supplementation and Childhood Morbidity from D and Respiratory Infections: A Meta-analysis." *Journal of Pediatrics* 142, no. 3 (March): 304.
- Haggerty, P. A., et al. 1994. "Community-Based Hygiene Education to Reduce Diarrhoeal I in Rural Zaire: Impact of the Intervention on Diarrhoeal Morbidity." *International Journal demiology* 23, no. 5 (October): 1050–1059.
- Hoke, C. H., et al. 1988. "Protection against Japanese Encephalitis by Inactivated Vaccines *England Journal of Medicine* 319, no. 10 (September): 608–614.
- IFDC. 2001. *Monthly Bulletin of the International Institute of Soil Fertility Management (Africa)* 14:2 (February).
- Jalan, Jyotsna, and Martin Ravallion. 1998. "Are There Dynamic Gains from a Poor-Area opment Program?" *Journal of Public Economics* 67:1 (January): 65–85.
- Jha, P., et al. 2001. "The Evidence Base for Interventions to Prevent HIV Infection in L Middle-Income Countries." Commission on Macroeconomics and Health, Working Gr Working Paper 2.
- Kaboski, Joseph, and Robert Townsend. 2002. "Policies and Impact: An Analysis of Villag Microfinance Institutions." Mimeo, University of Chicago.
- Knight, M., et al. 2000. "Antiplatelet Agents for Preventing and Treating Pre-eclampsia (C Review)." In *The Cochrane Library*, February 2000:2. Oxford: Update software, 2000.
- Kremer, Michael. 2003. "Randomized Evaluations of Educational Programs in Developing tries: Some Lessons." *American Economic Review* 93, no. 2 (May): 102–106.
- Kremer, Michael, and Edward Miguel. 2004. "Worms: Identifying Impacts on Educati Health in the Presence of Treatment Externalities." *Econometrica* 72, no. 1 (January): 159–.
- Kremer, Michael, Sylvie Moulin, and Robert Namunyu. 2003. "Decentralization: A Cau Tale." Mimeo, Harvard University.
- Mahomed, K. 1997. "Folate Supplementation in Pregnancy (Cochrane Review)." In *The C Library*, August. Oxford: Update Software.
- Meek, S., J. Hill, and J. Webster. 2001. "The Evidence Base for Interventions to Reduce M in Low and Middle-Income Countries." Commission on Macroeconomics and Health, W Group 5, Working Paper 6.
- Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Health and tion in the Presence of Treatment Externalities." *Econometrica* 72, no. 1:159–217.
- Morduch, Jonathan. 1998. "Does Microfinance Really Help the Poor? New Evidence from ship Programs in Bangladesh." Research Program in Development Studies. Mimeo, Pr. University.
- Morra, Linda G., and Ulrich R. W. Thumm. 1997. *1995 Evaluation Results*. Washington Operations Evaluation Department. World Bank.
- Munjanja, S. P., G. Lindmark, and L. Nystrom. 1996. "Randomized Control of a Reduce Programme of Antenatal Care in Harare, Zimbabwe." *Lancet* 348, no. 9034 (August): 364–

- Mwinda, A., et al. 1998. "Twice Weekly Tuberculosis Preventive Therapy in HIV Infection in Zambia." *AIDS* 12, no. 18 (December): 2447-2457.
- Narayanan, Deepa, ed. 2000. *Empowerment and Poverty Reduction: A Sourcebook*. Washington, D.C.: World Bank.
- O'Brien, Katherine L., et al. 2003. "Efficacy and Safety of Seven-Valent Conjugate Pneumococcal Vaccine in American Indian Children: Group Randomized Trial." *Lancet* 362, no. 9381 (August): 355-361.
- Pérez-Schael, Irene, et al. 1997. "Efficacy of the Rhesus Rotavirus-Based Quadrivalent Vaccine in Infants and Young Children in Venezuela." *New England Journal of Medicine* 337, no. 17 (October): 1181-1189.
- Pritchett, Lant. 2002. "It Pays to be Ignorant: A Simply Political Economy of Rigorous Program Evaluation." *Policy Reform* 5, no. 4:251-269.
- Roberts, Les, et al. 2001. "Keeping Clean in a Malawi Refugee Camp: A Randomized Intervention Trial." *Bulletin of the World Health Organization* 79, no. 4:280-287.
- Rowland, M., et al. 2000. "Indoor Residual Spraying with Alphacypermethrin Controls Malaria in Pakistan: A Community-Randomized Trial." *Tropical Medicine and International Health* 5, no. 7:472-481.
- Rush, D. 2000. "Nutrition and Maternal Mortality in the Developing World." *American Journal of Clinical Nutrition* 72 (suppl): 212S-240S.
- Schultz, Paul. 2001. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." Economic Growth Center, Center Discussion Paper 834.
- Shaffer, N., et al. 1999. "Short-Course Zidovudine for Perinatal HIV-1 Transmission in Bangkok, Thailand: A Randomized Controlled Trial." *Lancet* 353, no. 9155 (March): 773-780.
- Shulman, C. E., et al. 1998. "A Community Randomized Controlled Trial of Insecticide-Treated Bednets for the Prevention of Malaria and Anaemia among Primigravid Women on the Kenyan Coast." *Tropical Medicine and Internal Health* 3, no. 3 (March): 197-204.
- Stanton, B. F. 1987. "An Educational Intervention for Altering Water-Sanitation Behaviors to Reduce Childhood Diarrhea in Urban Bangladesh. II. A Randomized Trial to Assess the Impact of the Intervention on Hygienic Behaviors and Rates of Diarrhea." *American Journal of Epidemiology* 125, no. 2 (February): 292-301.
- Tinker, Anne, Kathleen Finn, and Joanne Epp. 2000. "Improving Women's Health: Issues and Interventions." Washington, D.C.: World Bank.
- Tuberculosis Research Center (ICMR), Chennai. 1999. "Fifteen Year Follow Up of Trial of BCG Vaccines in South India for Tuberculosis Prevention." *Indian Journal of Medical Research* 110 (August): 56-69.
- Vaccine Fund. 2002. *Strategic Plan 2002-2006*. Washington, D.C.: The Vaccine Fund.
- van de Walle, Dominique. "Impact Evaluation of a Rural Road Rehabilitation Project in Viet Nam." Cited 11 February 2004. Available online (<http://econ.worldbank.org/view.php?type=20&id=11865>).
- Verlé, P., et al. 1999. "Control of Malaria Vectors: Cost Analysis in a Province in Northern Vietnam." *Tropical Medicine & International Health* 4, no. 2 (February): 139-145.
- Vermeersch, Christel. 2002. "School Meals, Educational Achievement and School Competition: Evidence from a Randomized Experiment." Mimeo, Harvard University.
- Villar, J., et al. 2001. "WHO Antenatal Care Randomized Trial for the Evaluation of a New Model of Routine Antenatal Care." *Lancet* 357, no. 9268 (November): 1551-1564.
- Walker, Damien, et al. 2002. "An Economic Analysis of Midwifery Training Programmes in South Kalimantan, Indonesia." *Bulletin of the World Health Organization* 80, no. 1:47-55.
- Wawer, M. J., et al. 1999. "Control of Sexually Transmitted Diseases for AIDS Prevention in Uganda: A Randomized Community Trial." Rakai Project Study Group. *Lancet* 353, no. 9152 (February): 525-535.

- Wichmann, W., eds. n.d. *World Fertilizer Use Manual*, available online at (<http://www.fifa.org/ifa/publicat/html/pubman/manual.htm>), International Fertilizer Industry Association.
- Wiktor, S. Z., et al. 1999. "Short-Course Oral Zidovudine for Prevention of Mother-to-Child Transmission of HIV-1 in Abidjan, Côte d'Ivoire." *Lancet* 353, no. 9155 (March): 781–785.
- World Bank. 1989. *Project Performance Results for 1987*. Washington, D.C.: Operations and Policy Evaluation Department, World Bank.
- World Bank. 1993. *Evaluation results for 1991*. Washington, D.C.: Operations and Policy Evaluation Department, World Bank.
- World Bank. 1994. *The World Bank Annual Report 1994*. Washington, D.C.: World Bank.
- World Bank. 2000a. *1999 Annual Review of Development Effectiveness*. Washington, D.C.: Operations and Policy Evaluation Department, World Bank.
- World Bank. 2000b. *1999 World Bank Annual Report*. Washington, D.C.: World Bank.
- World Bank. 2001. *2000 World Bank Annual Report*. Washington, D.C.: World Bank.
- World Bank. 2002a. *2001 Annual Review of Development Effectiveness*. Washington, D.C.: Operations and Policy Evaluation Department, World Bank.
- World Bank. 2002b. *2001 World Bank Annual Report*. Washington, D.C.: World Bank.
- World Bank. 2003a. *2002 Annual Review of Development Effectiveness*. Washington, D.C.: Operations and Policy Evaluation Department, World Bank.
- World Bank. 2003b. *2002 World Bank Annual Report*. Washington, D.C.: World Bank.