The Future of Aid: Building Knowledge Collectively

Ruth Levine and William Savedoff

Abstract

This paper articulates how development assistance can promote program evaluation generally, and impact evaluation specifically, as a contribution to good governance. We argue that aid agencies are particularly well suited to fund impact evaluations, and can accelerate progress in the developing world by increasing the resources available for evaluation, particularly through a collective vehicle like the International Initiative for Impact Evaluation (3ie). Finally, we highlight the conditions that need to be in place – and require additional efforts – to yield the full benefits of collective investment in finding out what works.

The authors would like to thank Albert Alwang for his research assistance; Drew Cameron and Anjini Mishra for sharing pre-publication drafts of their paper; and three anonymous reviewers who enriched the paper. All remaining errors are the authors' sole responsibility. An earlier draft of this working paper was prepared for a festschrift in honor of the tenure of Howard White as Executive Director of the International Initiative for Impact Evaluation. The authors were instrumental in the creation of the International Initiative for Impact Evaluation (3ie) in 2009 through their work with CGD's "Closing the Evaluation Gap" initiative. However, neither author has received funding from 3ie or been involved in its management since 2010.

Ruth Levine is Director, Global Development and Population Program, William and Flora Hewlett Foundation. Contact: rlevine@hewlett.org. William Savedoff is a senior fellow at the Center for Global Development. Contact: wsavedoff@cgdev.org.

Ruth Levine and William Savedoff. 2015. "The Future of Aid: Building Knowledge Collectively." CGD Policy Paper 50. Washington DC: Center for Global Development.
http://www.cgdev.org/publication/future-aid-building-knowledge-collectively

Center for Global Development
2055 L Street
Fifth Floor
Washington DC 20036
202-416-4000
www.cgdev.org

Center for Global Development
Independent research and practical ideas for global prosperity

CGD Policy Paper 050
January 2015

Contents

Introduction

All governments face the same problem: how can they know whether the actions they take to benefit citizens are successful or are, instead, wasting tax dollars and slowing social and economic progress? Obtaining that knowledge is hard and often considered a quixotic ambition, particularly in the data-poor environments of many middle- and low-income countries. Taking time to learn how well government programs work has also been criticized as a technocratic sideshow to the main stage of politics.

The tide is turning, however. Throughout the world policymakers and citizens alike are recognizing that the very legitimacy of public sector institutions is jeopardized by their inability to demonstrate the positive differences they make and, when necessary, to change course to improve performance. Politicians are increasingly demanding "value for money," citizens have the ability to quickly and widely broadcast complaints against the State, and standards of openness and accountability are trending upward. Evaluating and using evaluation results are increasingly seen as activities that are as intrinsic to good government as transparency.

While the evaluation of public policies and programs relies on innovations and experiences developed over more than half a century (Rossi et al. 2003, pp. 9-20), in recent years researchers and practitioners have greatly expanded the application of new methods to program evaluation in low- and middle-income countries, seeing this as a fundamental tool for social progress. Building on experience in industrialized countries, academic researchers, government officials, individuals at bilateral and multilateral agencies and non-governmental organizations have promulgated innovative evaluation approaches that are appropriate for varied contexts in middle- and low-income countries. Contemporary leaders in South Africa, Mexico, Colombia, Brazil, Indonesia, Rwanda, Kenya and many other countries have committed to evaluation as an instrument of accountability to voters, and a means of fulfilling their executive responsibilities. By interrogating the effectiveness of efforts to prevent disease, improve learning outcomes, increase family incomes, and reduce gender bias, supporters of program evaluation are contributing both to improvements in specific interventions and to the larger cause of enlightened social and economic policy.

In this paper, we seek to articulate how program evaluation generally, and impact evaluation specifically, contribute to good governance – not as a replacement for politics, but as means of both learning and accountability. We then argue that institutions with the mandate to

accelerate progress in the developing world through foreign aid[1] – aid agencies – are particularly well suited to fund impact evaluations. We argue, in fact, that funding policy-relevant impact evaluations through a collective vehicle like 3ie should be one of their *primary* activities. Finally, we highlight the conditions that need to be in place – and require additional efforts – to yield the full benefits of collective investment in finding out what works.

Politics First, Effectiveness Second

Core social choices are worked out in political processes, whether democratic or otherwise. Questions such as assigning priority to defending borders versus improving schools or building roads are answered through political negotiations that reflect collective values and power relationships. Despite efforts to override processes to arrive at a set of social choices – for example, by asserting a set of affirmative universal rights or by advocating "value-neutral" tools like cost-benefit analysis – government priorities are rightly established through the wonderful and messy human process referred to as "politics." Evidence, knowledge and technical expertise has its role to play in this process, but it is neither determinate nor sufficient. Rather evidence is itself contested in this forum but it does inform and shape debates.

Once these choices are made, the tasks facing governments are how to design, fund and execute often massive public programs that are aligned with those priorities, and then to measure progress against expectations. Governments have to sort out how to identify and reach target populations, how to set benefit levels, how to deliver services of high quality but affordable cost, and many other tricky issues for which there is no recipe or playbook. In the education sector, for example, one political administration may wish to expand the role of private providers while another may seek to universalize and improve public education. While the agendas differ, they both imply a need to figure out how to use public dollars and policies to achieve the goals. It is at these stages that technical, empirical tools have more direct benefit, influencing managerial choices, regulatory decisions, and policy design.

[1] In this paper, we use "foreign aid" to mean the financing provided by members of the OECD/Development Assistance Committee. This also includes grant-type funding from the World Bank and other multilateral development banks. While we do not explicitly discuss the features of funding from private foundations, most of the same arguments apply. We do not cover financing through loans provided by the World Bank or other multilateral development banks, although a case could be made that these could and should be used, in part, to support impact evaluation if other resources were not available.

While all of the technical tasks are difficult, perhaps the most difficult to undertake in a systematic and sustained manner is the measurement of progress. Yet without it, the public sector perpetually lacks the information required for improving program design; has difficulty sustaining support from constituents when opposition emerges; and finds implementation bottlenecks challenging to overcome.

The problem of measuring what matters, faced by governments of all countries, is particularly important to solve in middle- and low-income countries. With vastly more needs than domestic (plus donor) funding can meet, with weak and unreliable official statistics, and with severely limited technical capacity within government agencies, policy makers in developing countries typically operate in the dark. Yet the stakes are extraordinarily high. An inability to know what's working is very costly, resulting in scarce funding and political capital being wasted on ineffective if well-intentioned schemes.

Evaluation Holds Much Promise

In many developing countries, so little attention has typically been given to empirical information and technical considerations that the design or modification of health, education and anti-poverty programs is influenced by the latest ideas from consultants sent by donor agencies; by improvised adaptation of efforts in neighboring countries; or by guesswork. The opportunities for false assumptions and self-interest to affect program design and implementation are manifold.

Public officials are not the only ones who operate in the dark or on the basis of the limited signs of success or failure that they can observe directly. Citizens are similarly constrained. Other than public budget information – which is increasingly available to the public thanks to the "open budgets" movement – citizens and the groups that organize on their behalf have few sources of information about how well or poorly government programs are being implemented. They have almost no information about the effect of government programs on outcomes such as improvements in health within disadvantaged communities, reductions in sexual violence, improvements in the ability of school age children to read and write, increases in the income of women in poverty, or improvements in the productivity of small-scale farmers receiving seed, fertilizer and training. Without such information, they are lacking crucial facts that could inform their votes or citizen action.

This is where many types of program evaluation demonstrate their value. Program evaluation includes dispassionate assessment of whether a program was implemented as

designed. Rigorous factual analysis can detect how many seemingly well-designed programs lose their way in basic implementation (White 2009). This might include, for example, situations in which the beneficiaries are not identified well, the staff are poorly trained, or supplies are stuck at the port of entry. A central task of examining the effectiveness of government programs is to simply answer the question: Was the program implemented as designed? If not, why?

In Kenya, for example, a World Bank-financed project sought to improve agricultural extension practices, and yet the evaluation found little change in what extension agents were doing during the project lifetime; only 7 percent of participating farmers had the amount of contact with extension agents that the project design had anticipated. In Bangladesh, most of the women and children who were supposed to receive supplementary feeding in a large nutrition program did not. This type of execution failure is prevalent, and can be detected with basic program evaluation methods that track actions to see whether implementation occurred as planned (White 2009).

In addition to identifying execution failures (and successes), program evaluation can provide valuable information about the cost of interventions and targeting strategies and the system outputs (such as the number of trainees or the number of women with access to savings accounts). It can shed light on institutional strengths and weaknesses that influence the ultimate sustainability of any effort. It can reveal the meaning and interpretation of change as experienced by beneficiaries themselves.

Evaluations which assess execution, operations, costs, strategies, institutional development, and meaning all answer important questions. Another set of fundamental questions relates to impact in terms of outcomes. These questions are:

- Did the program, when implemented as designed, improve outcomes?
- Were the gains large enough to be worth the cost? and
- Are the gains larger than would have been produced with alternative ways of using the same resources?"

These questions, important as they are, are rarely answered. Each hinges on an ability to measure the net impact of a particular program on a defined set of outcomes at the individual and/or community level. Furthermore, the usefulness of answering these

questions for a particular program is limited unless situated within a larger body of evidence from which to assess the reliability of findings and compare the program with alternatives.

Impact Evaluation is an Evolving and Growing Field

In 2006, the Center for Global Development Evaluation Gap Working Group pointed out in its report, *When Will We Ever Learn: Improving Lives through Impact Evaluation* (2006), that evaluations of the impact of both government and donor-funded programs had been systematically underfunded despite their potential utility (Evaluation Gap Working Group 2006). The working group cited three main reasons for this: a classic "public goods" problem, in which the benefits of the investment by a few would accrue to many others; the imperative to get programs implemented rather than to take the time to build in evaluation; and the difficulty of large bureaucracies to take in and act on news about disappointing results.[2]

Although many of the reasons for underinvestment in high-quality, relevant impact evaluation persist, much has been achieved since 2006. Improvements include better methods, broader application across sectors and topics, accumulation of bodies of evidence rather than isolated studies, and the production of systematic reviews that have the potential to provide balanced guidance to the policy community.

The evaluation of impact is methodologically ambitious, because it requires estimating what would have been observed in the absence of an intervention, in addition to what is actually observed. Only rarely is it possible to judge the net effect of an intervention on outcomes based solely on data collected at a project's conclusion. Instead, researchers and practitioners try to compare outcomes before and after an intervention or analyze cross-sectional data that includes participants and non-participants. In certain contexts, these approaches are effective; but for a large range of programs they are unable to account for other factors that may be driving change. Some more advanced statistical methods that used to estimate the difference between what happened and what might otherwise have happened include instrumental variables, difference-in-differences, matching, and regression discontinuity. Qualitative studies which gather data from non-participating individuals or organizations are

[2] Other reasons proposed to explain the underinvestment in impact evaluations include that "it pays to be ignorant" (Pritchett 2002) and that charitable motives and national political interests are of greater significance than aid effectiveness to policy decisions (Gaarder and Bartsch 2014).

similarly engaged in extracting information that is used to account for confounding factors (Rossi et al. 2003).

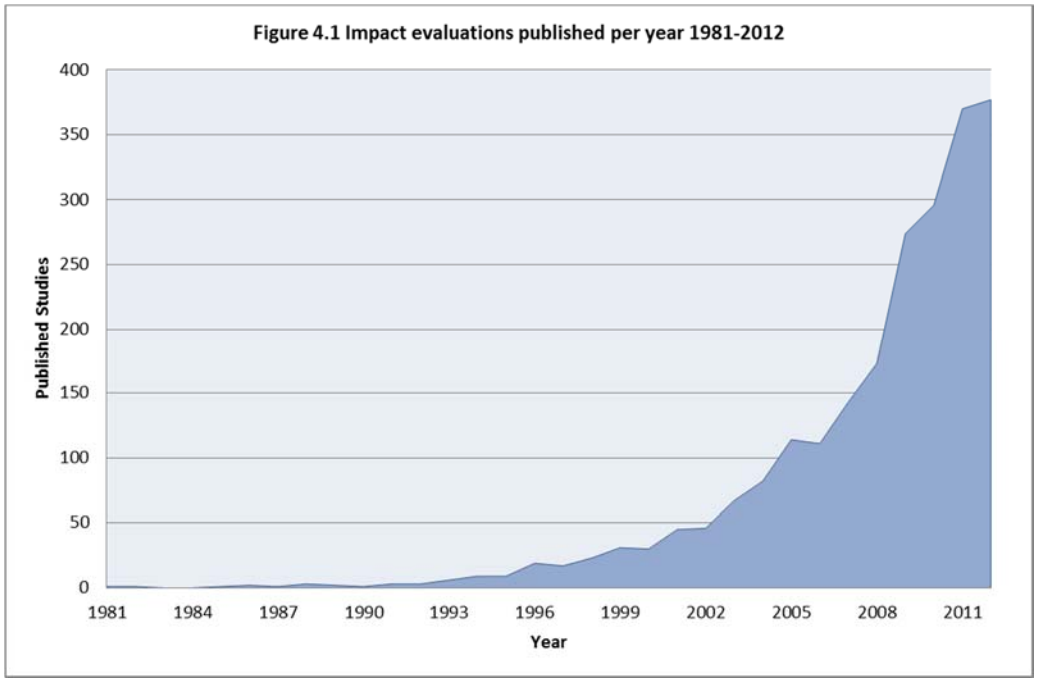
A large part of the recent growth in impact evaluation has been spurred by advances in using methods derived from research on the effectiveness of medical interventions. Development economists have adapted randomized control trial methods to compare health, education, behavioral and economic outcomes among those exposed to or participating in an intervention with outcomes in a similar group of individuals not exposed or participating (White 2013). Field trials, however, operate in a far messier environment than many clinical trials testing a drug against placebo or an alternative therapy. In response, the growing impact evaluation community has, over the past decade or so, developed creative yet rigorous ways to randomize the individuals or geographic areas to which a program is extended (De Allegri et al. 2008); integrated assessment of the quality or intensity of program implementation into the evaluation design (Garbarino et al. 2009); tested ways to shorten the time spent on evaluating impact (Cody and Asher 2014); conducted evaluations in the context of multiple demographic surveillance sites (LSHTM 1979-Present); compared findings from experimental and quasi-experimental methods (Bifulco 2012); and pursued "variation in treatment" rather than solely "with/without" designs, to yield more policy-relevant findings (Schochet et al. 2014). While the methodological challenges are far from being solved, impact evaluation has proven to be a field that innovates quickly, and adopts new solutions as they are developed.

Since 2006, the impact evaluation community has also demonstrated a boldness of ambition in its choice of subject matter. Early applications of impact evaluation focused primarily on health and education outcomes that had reasonably standard definitions and could be measured and compared at the individual level. Today, impact evaluators using both random assignment and non-experimental methods routinely tackle an extraordinarily broad array of topics: gender-based violence (Kiplesund and Morton 2014), road traffic accidents (Habyarimana and Jack 2009), forest preservation (Hatanga 2014), corruption (Zamboni and Litschig 2013), trauma following rape (Bass et al. 2013), savings behavior (Karlan and Goldberg 2007), income gains (Haushofer and Shapiro 2013), women's empowerment (Beaman et al. 2009-2011), teacher absenteeism, health worker performance and health outcomes (Basinga et al. 2011), accountability through citizen action (Results for Development 2013) and many more.

An important although hidden benefit of the forays by impact evaluators into domains in which the outcomes are very difficult to measure is the conceptual clarity they have fostered. Merely being engaged in an impact evaluation requires that, from the outset, program designers and implementers can clearly articulate what they are trying to achieve. Rather than being able to say that they are "addressing gender inequality," for example, they must be able to identify the hoped-for changes that can be directly observed or on which people can report in a valid and reliable way. The work of the World Bank's Gender Innovation Lab represents an example of the contribution impact evaluators can and do have on fostering intellectual discipline in fields that otherwise might lean toward expert judgment rather than reproducible measurement.

Beyond methods and scope, in recent years the impact evaluation community has simply been very busy doing impact evaluations. A forthcoming study from researchers at 3ie has the most complete database of published impact evaluations available and it finds that as recently as 1995, fewer than 10 studies of developing country policies were published each year (See Figure 1). That number has grown rapidly, with more than 300 studies coming out annually (Cameron and Mishra 2014). The Abdul Jameel Latif Poverty Action Lab (JPAL) and Innovations for Poverty Action (IPA) are relatively new research centers dedicated to conducting impact evaluations and in 2014 they listed 567 and 220 ongoing studies, respectively. Other research centers have substantially increased their impact evaluation work, whether older ones like the International Food Policy Research Institute or newer ones like the Peruvian think tank GRADE.

Figure 1. Impact Evaluations Published Per Year, 1981-2012



Source: Reproduced from Drew Cameron and Anjini Mishra. 2014.

Many agencies are involved in funding or commissioning this work. Since it started in 2009, 3ie has awarded 131 grants for impact evaluations (3ie 2014), 19 of which are now publicly available. The World Bank completed an average of 57 impact evaluations each year from 2005 to 2010 (IEG 2012). Just one of its initiatives, Development Impact Evaluation (DIME), has 131 studies in its working paper series of which 31 were added in 2013 (World Bank 2014). Spain contributed more than $13 million to the Spanish Investment Impact Fund (later renamed the Strategic Impact Evaluation Fund – SIEF) and the United Kingdom has contributed more than $40 million to impact evaluation work just on the basis of contributions to SIEF and 3ie. The Bill & Melinda Gates Foundation contributed about $45 million to 3ie from 2009 to 2014 in addition to other similar research that it supports through direct contracting or other institutions. Organizations as varied as Care, the International Rescue Committee, the Nike Foundation, the Inter-American Development Bank, the US Agency for International Development and the Asian Development Bank are implementing plans to increase support for and use of impact evaluations. This represents a substantial growth in funding for impact evaluations – perhaps as much as US$50 million a year – but it is still extremely small relative to the range of programs in developing countries which are being financed through foreign aid (more than US$100 billion each year) and through domestic developing country budgets, which are measured in trillions of dollars.

Developing countries themselves are more engaged in impact evaluation work than ever before. Countries like Mexico, Colombia, Chile, South Africa, and India have created dedicated government units concerned with evaluating public programs and commissioning or conducting a growing number of impact evaluations. Others like Uganda, Rwanda, Kenya and the Philippines are still actively seeking evidence to guide their policies through commissioning research or collaborating with international research teams.

With the accumulation of impact evaluations comes the opportunity to undertake systematic reviews and meta-analyses. These rigorous reviews of impact evaluations covering similar interventions go beyond isolated findings that are useful for a specific situation to provide an overview of the evidence across different contexts, insights regarding which theories are more useful, and opportunities to assess generalizability (White and Waddington 2012; Waddington et al. 2012). 3ie itself has produced 12 systematic reviews (3ie 2012), on topics ranging from the impact of field schools on agriculture productivity to interventions to prevent HIV through behavior change. The International Development Coordinating Group of the Campbell Collaboration began to publish systematic reviews in 2012 and now has 40 studies on its website.

As is true for all research related to social systems, the findings from impact evaluations produce concentric circles of benefit. First and foremost, the findings can be used by the agency or other organization implementing a program. If the evaluation demonstrates that meaningful improvements are being achieved, commensurate with the cost, the agency has information to sustain and expand the program. If results are disappointing, it can modify the design or take a new approach altogether. The Indian education non-governmental organization Pratham, for example, works collaboratively with JPAL to test and then refine a range of interventions, from those intended to reduce teacher absenteeism to different ways to teach reading and math skills.[3] The International Rescue Committee uses impact evaluation methods to test and then refine many of its interventions, sharing insights with other humanitarian organizations (Goldstein and McKenzie 2013).

Beyond those involved with the program being evaluated, others working with similar problems and populations can benefit as well. Facing similar problems, this next ring of people can build programs around interventions that have shown success or at least avoid mistakes made by others. In the United States, for example, the Department of Health and

[3] http://www.povertyactionlab.org/partners/pratham accessed Oct. 27, 2014.

Human Services dedicates funding to encourage replication of teen pregnancy prevention interventions, such as particular types of sex education, that have been demonstrated to be effective in rigorous impact evaluations (Office Of Adolescent Health 2014); a strategy that is increasingly used by the government on many social policies (Haskins and Margolis 2014). In the context of the developing world, the diffusion of the innovation of conditional cash transfers has been greatly facilitated by a series of impact evaluations, whose results have been aggregated to draw out inferences about what this type of anti-poverty program can achieve, and under what conditions (Davis et al. 2012). It was, in fact, the evaluation of the PROGRESA program in Mexico that inspired Mayor Michael Bloomberg to implement and evaluate a cash transfer program to incentivize families to maintain good school attendance in New York City (Bosman 2010). Pioneering work in assessing learning in India through a simple test of literacy and numeracy (ASER) test has inspired similar assessment initiatives in other countries, including Uganda, Kenya and Tanzania (Uwezo).

At the outermost ring of concentric benefits, and most removed from the original program's context, impact evaluations contribute to the body of evidence that helps to steer funders toward one set of approaches and away from others. Systematic reviews can serve as a point of departure for advancing a field, rather than repeating errors inadvertently (Waddington et al. 2012). For example, a systematic review of nine impact evaluations of programs aimed at improving teacher attendance in schools in developing countries found that "A combination of better monitoring and powerful incentives is effective in tackling teacher absenteeism. However, having a teacher in the classroom does not appear to be sufficient to improve student achievement (Guerrero et al. 2012)." This summary finding helps public sector officials in the education sector in two ways: First, it provides a basis for context-specific experimentation with combinations of enhanced supervision and strong financial incentives – not a blueprint for action, but a starting point that is more likely than guesswork to result in good outcomes. Second, it provides a caution against viewing teacher absenteeism as the only problem to solve before learning outcomes will improve.

Future Progress in Impact Evaluation is Threatened

Despite the dynamism of the field of impact evaluation, and both the realized and the potential benefits, future progress is threatened by several forces. First and foremost is what Lant Pritchett and others have referred to as the "hype cycle" (Pritchett 2013). In the face of real-world constraints and delays, overenthusiasm about what impact evaluation can achieve, and/or unrealistic expectations about how quickly policymakers will take up the findings

from evaluations, may lead to disillusionment. Second is the continued conflicts among professional evaluators and researchers around methods, and particularly the use of randomization to permit causal inferences to be drawn. Remarkably, a sterile and unproductive debate continues around methodological choices, generating far more heat than light.[4]

Third, those with the skills to do impact evaluations often pursue different interests than those who are responsible for policy decisions. Researchers have incentives to do studies that can be published in prestigious journals and enhance their standing among their peers. However, studies that replicate existing studies in order to assess the reliability or generalizability of findings are not valued. Furthermore, evaluators do not always find the questions being asked by policymakers to be interesting or researchable. Consequently, not all impact evaluations have clear benefits beyond being a contribution to the research literature.

Finally, relatively few public or private funders have invested in impact evaluation. Although significantly more funding is available today than the last decade, multiple institutions are seeking support from the same small pool of committed funders to advance their impact evaluation work. The field depends in an unhealthy way on the sustained interest of those funders. Unless a broader range of governments and organizations can be convinced of the need for collectively providing long-term stable funding to this knowledge-building endeavor, enormous opportunities will be lost.

Various solutions may be found to address these threats and problems but fundamental to almost all of them is maintaining and increasing financial support, particularly through a collective, fit-for-purpose mechanism like the 3ie. That is where an increased commitment by foreign aid to advancing evaluation as an essential element of good governance enters the picture.

Aid is Uniquely Suited to Impact Evaluations

Foreign aid can be helpful in many ways, but it is most useful for learning how to make public programs more effective. In this case, we are referring to public programs financed

[4] Examples of the debate include Development Channel Staff 2012; Villamor 2014; and Savedoff 2014.

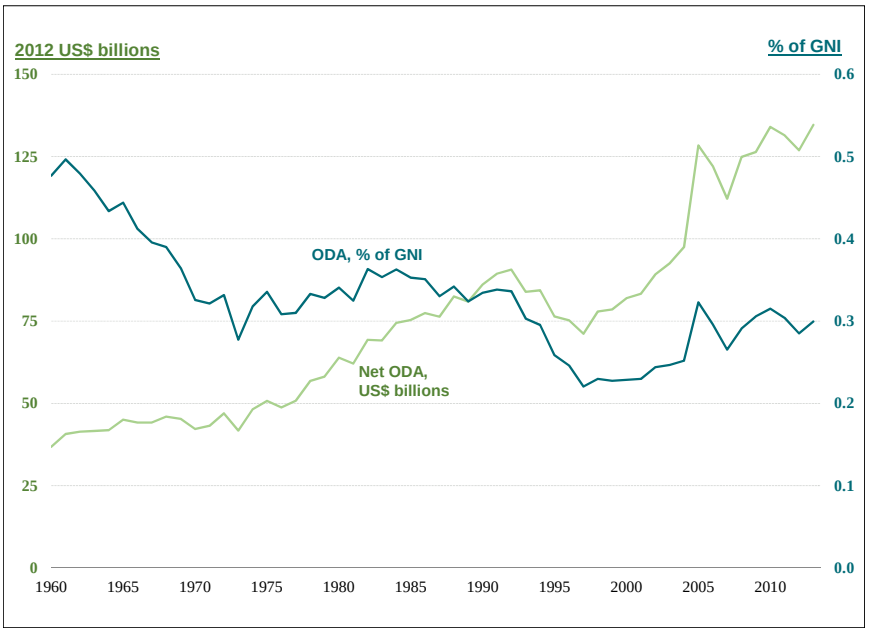
with any combination of domestic and external resources and operated by government agencies at either the national or subnational level or by NGOs.

Aid is uniquely suited to financing the impact evaluations that provide strong evidence and policy-relevant knowledge that can benefit many, and that can help build and reinforce trust between governments and citizens. Aid is uniquely suited for this task because of its small relative size as domestic finance grows; its ambition of disproportionate influence; its sensitivity to being used for illicit purposes; its ability to bridge several communities; and its aspirational role in advancing public sector accountability.

At one time aid served to fill a financing gap that held countries back, providing the resources that would otherwise not be available for large infrastructure and energy projects, manufacturing investments, and later to construct and supply schools and health facilities. This is no longer the case because over the past 20 years most low- and middle-income countries are increasingly attractive to private investors and governments have experienced rapid growth in their own revenues. Overseas development assistance has fallen as a share of donor country Gross National Income (GNI) from 0.5 percent in the 1960s to about 0.3 percent in this decade (see Figure 2). After levelling off in the 1990s, the absolute value of official aid flows started rising but plateaued again after the financial crash of 2007-2008. OECD countries have disbursed about US$125 billion per year since 2005. Chinese transfers have been rising over the decade but not enough to offset the relative decline from OECD countries.[5] At the same time, aid is declining in significance relative to the national income of receiving countries. Aid flows have fallen since the 1990s to an average of 12 percent of GNI in low-income countries (a shrinking category) and represent about 3 percent of GNI in middle-income countries that are still receiving aid (See Figure 3).

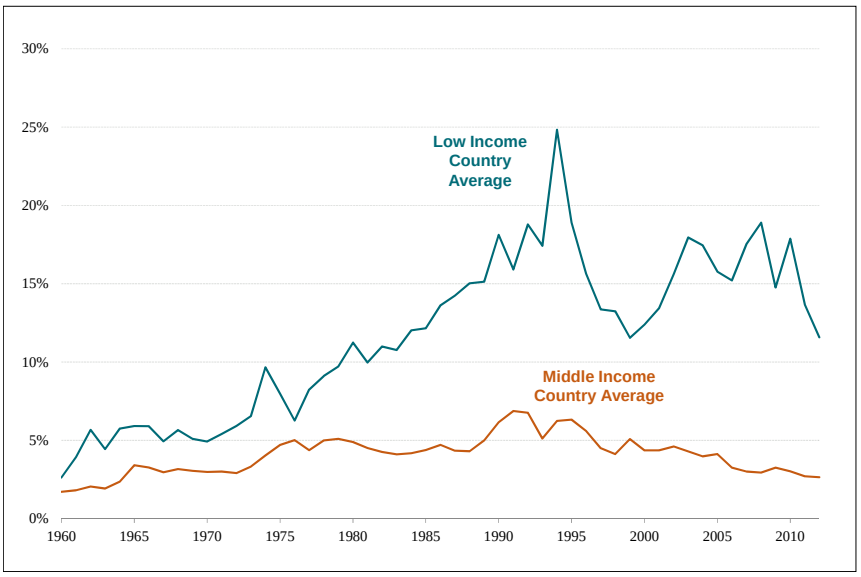
[5] Strange et al. 2013, estimated that Chinese transfers to 50 African countries rose from about US$2.8 billion in 2000 to about US$9.8 billion in 2010.

Figure 2: Official Development Assistance, 1960-2012



Source: OECD.Stat

Figure 3: Official Development Assistance as share of recipient country GNI (%)



Notes: The figure shows the average ratio of official development assistance (ODA) to Gross National Income (GNI) for countries within each income group. Countries were include if: they had a population greater than 1 million; received ODA; and have GNI reported in the database. Therefore the number of included countries varies from year to year.

Source: World Bank Development Indicators

The erosion in the value of aid as a source of financing for priority government programs is a positive development because aid can be volatile and disruptive. From the perspective of any given country, aid is highly unpredictable over time. Political choices in donor countries are influenced by changing geopolitical interests, sectoral focus, and fiscal conditions, yet they dictate how much aid is provided, to whom and for what (Desai and Kharas 2010). The resulting volatility and uncertainty creates problems for planning and continuity of public programs (Kharas 2008). For example, the United States decision to begin phasing out its support for AIDS treatment in South Africa and other countries demonstrates how foreign aid can fill gaps but simultaneously disrupt the domestic political process of negotiating priorities and developing locally appropriate, affordable strategies for vital national challenges (Kavanagh 2014).

Nevertheless, aid still has an important role to play because the problems it tackles – such as poor health outcomes, low educational attainment, and low agricultural productivity – are still with us even after years of impressive gains. Not only do these problems persist in many parts of the world, we now have new challenges, such as climate disruption and the rise of non-communicable diseases. In addition, now that many "bricks-and-mortar" solutions are in place, we collectively conceptualize the causes of poor development outcomes differently. Increasingly, we recognize the role of governance, social norms and incentives in shaping outcomes at the individual and societal level (Grindle 2004; World Bank 2004; World Bank 2015).

Recognizing the relative decline in financing capacity and understanding the underlying causes of persistent problems in new ways, aid agencies are trying to make the most of the resources available to them. One path is to increasingly concentrate on the small and shrinking set of very poor countries in which aid still represents a large share of resources. Another path is to specialize in humanitarian response and work in post-conflict environments where immediate needs, the absence of private sector investment, and minimally functional governments create a vacuum that aid agencies can partially fill. A third path is to contribute with knowledge. While this third path might be dismissed as "merely doing research," knowledge from such research has the potential to be the greatest source of sustainable benefits to which aid agencies can contribute.

As countries develop the public institutions that provide key services and support for growth, filling the gap in money becomes much less important than know-how. For example, a country like Argentina, with GDP per capita of more than US$14,000 and health

spending of almost $1,000 per capita does not need international agencies to fund its health system. Yet Argentina sought a World Bank loan for extending provincial health insurance to women and children. By borrowing from the World Bank, Argentina received technical assistance on a complex arrangement that altered incentives for national and provincial institutions, local healthcare providers, and beneficiaries. It also received support to do a rigorous impact evaluation of the program (Gertler and Giavagnoli 2014). As a result, Argentina knows that this program, which represented a mere 1% of the country's public health spending had significant impacts on the health of its citizens. Argentina also has the assurance that public policies instituted by the program will continue to generate benefits beyond the specific loan. By generating knowledge with relatively few dollars, the project leveraged resources far in excess of anything aid organizations could have directly offered or sustained.

A main advantage of foreign aid is its ability to mobilize external technical experts and bring them to collaborate with domestic researchers and evaluators. For a low-income country to engage an international expert to do impact evaluations can often cost ten times more than contracting domestically. And yet, the supply of domestic researchers with relevant technical skills and experience is often limited. Governments find it difficult to justify such a cost difference to the detriment of a study's quality, while international agencies using foreign assistance funds face fewer such constraints. Beyond the difference in costs, foreign funding is itself associated with international networks that can help identify, mobilize and engage the right technical experts. Since its founding, 3ie itself has assumed such a role. The combination of foreign funding and multilateral participation has allowed 3ie to generate a database of technical experts who can be called upon to review grant proposals, advise domestic researchers, and directly collaborate on impact evaluations.

Aid is not just suited to funding impact evaluations because of its size and ability to mobilize needed expertise. It is also suited to funding impact evaluations because doing so can make a significant contribution to better governance and greater responsiveness of governments to citizens. Instead of being in tension with democratic processes, as sometimes is the case when aid distorts domestic priorities, aid for impact evaluations of public programs can help provide crucial information to both governments and citizens that reinforce a healthy relationship.

We recognize that effectiveness is not the primary motivation for foreign aid. Countries choose to give foreign assistance for many reasons – demonstrating concern for less

fortunate people, getting national credit for action, as well as geopolitical and financial interests all play a role. But a large part of the debates over the uses and application of foreign aid occur within bureaucracies where effectiveness and impact are a visible and prominent concern. In the broader political and the more specific bureaucratic realms, evidence about effectiveness from impact evaluations therefore plays a role in informing and influencing choices.

The future of foreign aid has little to do with filling financing gaps for developing countries. Rather it is going to be increasingly focused on financing public goods, humanitarian assistance, and building knowledge. In building knowledge, foreign aid will be able to contribute significantly to development by supporting initiatives that systematically study public programs – whether those conducted by developing countries themselves or supported as pilot experiences by aid agencies. Foreign funding is better suited to financing studies – which complement domestic programs –than financing investment and services – that alter and substitute for domestic funding. It can mobilize the best technical expertise for collaborating with local researchers and it is driven, politically, to seek ways of leveraging positive benefits which can be achieved by revealing more effective instruments of public policy. To the extent it fosters the development of domestic institutions and capacity to research, assess, and learn about public programs, it can have even longer lasting effects.

Collective Is More Effective

Though bilateral investments in impact evaluation are helpful, collective action to fund impact evaluations is even more likely to succeed at advancing our knowledge about effective public programs. This is the case for several reasons. First, the knowledge generated by impact evaluations is a public good that provides insufficient incentive for any single actor to invest adequately relative to the benefits to everyone. Only collectively can we ensure adequate investment. Secondly, collective action can promote better quality studies from which we can have better and more reliable information. Finally, collective funding creates opportunities for efficiencies of scope and scale in the impact evaluation endeavor.

The benefit of an impact evaluation is the knowledge it provides. Such knowledge is a classic public good in the sense that one person can use it without using it up.[6] One government

[6] This feature is called "nonrivalry in consumption" in the economics literature. A second feature – non-excludability – is also required to characterize a pure public good. This distinguishes public goods from so-called "club goods" which are non-rivalrous in consumption but from which people can be excluded – such as cinemas

can learn from it without reducing its value for public accountability or for informing policies in other places. This is what makes knowledge such a powerful force for progress. It is the gift that keeps giving.[7]

This public good characteristic of knowledge, however, also means that if each individual, organization or country only invests in studies in proportion to the benefits that they themselves receive, then aggregate investment in building knowledge will be too low. Some spillovers will occur but not the full concentric benefits that would come from collective action to invest adequately. Bilateral aid programs are subject to this same limitation. They regularly face pressures to generate information that is useful to *their* programs and today's policy questions. In so doing, they forego opportunities to invest in studying programs in other places that might have had direct bearing on decisions today or which could yield valuable insights for decisions tomorrow.

The easiest course of action for governments giving or receiving aid is to be "free riders" – to rely on others to invest in research and take advantage of the resulting knowledge. This is a perfectly rational strategy from each country's perspective but clearly a losing proposition for the world as a whole. The best way to resolve such a problem is to change each government's incentives – for example by only giving them credit for aid programs that actually achieve outcomes (Gaarder and Bartsch 2014). An alternative solution to this free rider problem is to create a commitment device, that is, to find ways for all countries – or at least a significant group – to establish a visible, enforceable rule for financing impact evaluations. Such a rule might involve committing each country to finance a set share of its own programs – similar to international commitments to contribute 0.7% of GDP to foreign assistance or reduce domestic greenhouse gas emissions. An alternative is for countries to contribute set amounts to a multilateral institution as they do, for example, under agreements that fund the IMF or UN agencies. By collectively committing funds to impact evaluation studies, countries could shift from a situation with inadequate investment in building knowledge to one which comes closer to achieving the full concentric benefits that are possible. This was actually the key argument for proposing the creation of 3ie. Yet funding

---

and private parks. Public policies sometimes turn knowledge into an excludable good by creating patents and copyrights.

[7] One of the most dramatic examples of how knowledge has affected human wellbeing is the rise in human life expectancy. Life expectancy in the United States and Europe in the 1950s averaged around 68 years at income levels of 1,916 (PPP Converted GDP Per Capita, G-K method, at current prices I$). Countries with those same income levels today have life expectancies that are 20 years longer – due primarily to knowledge of healthier behaviors and public investment in cost-effective public health measures (Kenny 2012).

for 3ie in its first six years, which remains dominated by a small number of organizations, shows that this free rider issue is still a problem. Just two contributors – the Bill & Melinda Gates Foundation and the British government – accounted for US$30.8 million of 3ie's $31.7 million income in 2013 (3ie Annual Report 2013).

The collective decision to adequately fund impact evaluations may be hard, but creating an institution to receive and apply those funds has the ancillary benefits of improving the quality and usability of information. Institutions that fund research, like the US National Academies of Science, the Research Councils in many Western and Northern European countries, or Brazil's Oswaldo Cruz Foundation improve the quality of studies by formalizing peer review processes and insulating grant decisions from political manipulation. Their application procedures and grant decisions generate explicit and implicit standards for the academic and scientific communities that seek their funding. By working internationally, an institution like 3ie has these advantages and more. Its decisions are even less influenced by particular constituencies within particular countries and can take advantage of a larger community of disinterested parties with technical expertise.

A collective approach to funding impact evaluation is also efficient. First, grants can be directed toward clusters of studies that improve the reliability and generalizability of findings. The results of a single study might be incorrect for any number of reasons. By conducting several studies on similar policies in different contexts, it becomes possible to assess whether a particular finding is biased, mistaken, or idiosyncratic.[8] Having a number of studies with similar findings gives policymakers and researchers greater confidence in the conclusions.

Systematic reviews are *post facto* efforts to find and draw conclusions from studies that have a similar focus. Such systematic reviews put specific studies in the context of a wider body of knowledge and pay explicit attention to their methodological rigor (White and Waddington 2012). Public officials will only sustain interest in impact evaluations if they produce usable information – and systematic reviews are one way to demonstrate the usefulness of impact evaluation findings. Clustering studies explicitly around important questions is something a collective initiative can more easily undertake, pushing the frontier of knowledge more

[8]Some examples of clustered studies are reported in Mejía 2014 with regard to introducing laptops in schools; and Davis et al. 2012 regarding conditional cash transfer programs in Sub-Saharan Africa. MCC (n.d.) benefited from commissioning five comparable studies on agricultural extension services. 3ie has initiated a number of clustered studies on topics like social protection (see http://www.3ieimpact.org/en/funding/thematic-window/social-protection/award-winners/).

quickly, and avoiding the all too common experience of undertaking systematic reviews only to find that too few studies are available from which to draw firm conclusions.[9]

A collective institution is also more efficient than individual or bilateral initiatives because of scale economies in administering a grant program. Most bilateral agencies are too small to have staff with the expertise and time to dedicate to drafting requests for proposals, soliciting proposals, convening expert review panels, supervising grantees, and monitoring the quality of research. The marginal cost of reviewing an additional grant is small once a group of experts has been convened. Conversely, a grant program cannot generally afford to convene enough people with sufficiently specialized knowledge to assess proposals unless it is reviewing a significant number of applications.

A third advantage for a collective institution to undertake grant review and approval of impact evaluations is that it fosters cross-sector and cross-disciplinary learning. Methodological innovations in medicine have made their way into social policy research, econometric techniques have influenced education research, and mixed methods researchers are increasingly contributing a nuanced understanding of the "why" to the core impact evaluation question of "what changed and by how much." This kind of diffusion occurs more effectively and quickly when the community of researchers from different fields and sectors are brought together through a grant-making institution that fosters such communication. Additional learning also occurs when phenomena observed in one sector have relevance to another, such as when public service delivery issues in education arise in water or health or when poverty-reduction strategies in microfinance have a bearing on small business development.

A collectively-financed international institution for promoting impact evaluation is not a panacea by any means. Impact evaluations are always going to be subject to concerns that their findings may not be generalizable and that countries may be less likely to absorb lessons from studies financed, let alone conducted, by other countries. Collective financing of impact evaluations which are commissioned through an independent organization runs

[9] A systematic review by Gosden et al. 2001 found only six rigorous studies of payment systems on physician behavior, but a more recent systematic review (Van Herck et al. 2010) found 50 studies with concurrent comparisons or interrupted time series designs. Recognizing this as a major question facing health systems around the world, an international organization could finance a cluster of studies on this topic of use to everyone. UK and Norwegian funding for the World Bank's Health Results Innovation Trust Fund is an example in relation to this specific topic. 3ie has also clustered studies by issuing requests for proposals on specific topics such as HIV self-testing, climate change, and social protection (see "Thematic Windows" at www.3ieimpact.org).

the risk of losing opportunities to improve projects when evaluators can engage with project designers and implementers (Jacquet 2005; Gaarder and Bartsch 2014). The relevance of impact evaluations is also a regular concern, one which requires the institutions commissioning impact evaluations, whether bilateral agencies or a multilateral institute, to be open to an exchange of views between policymakers, managers, implementers, researchers, and beneficiaries about which questions are important to public policy decisions.

The question raised by such critiques is not whether impact evaluations should be collectively funded but whether collectively-funded impact evaluations can be conducted in ways that address these concerns by improving their generalizability and relevance. We already have a number of ways to address this latter question, many of them pioneered by bilateral and multilateral agencies or non-profit institutions like 3ie. These include:

a. Mechanisms to promote "practice relevant" impact evaluations, such as 3ie's policy window
b. Clustering studies around common questions so evidence from different contexts can be used to assess external validity, such as 3ies calls for proposals around specific issue areas
c. Programs to encourage domestic evaluation institutions such as Mexico's National Evaluation Council (CONEVAL) and improve local research capacity so that external evidence can be appropriately considered and domestic studies can be more rigorous
d. Contributions from developing countries to the collective institution so as to promote a sense of ownership and engagement, as well as authentic involvement in governance, guidance and debates, such as the engagement by Pakistan, South Africa, Uganda and other member countries in 3ie
e. Efforts to promote involvement of domestic researchers and policy evaluation groups in producing impact evaluations, as 3ie has attempted with its preference for local researchers and for local researchers' substantive engagement
f. Information exchanges between countries, as 3ie has done in international workshops and conferences

Under the right circumstances, aid is uniquely suited to financing impact evaluations and the most effective approach is to contribute to a collective initiative. Directing aid toward one or more international institutions can provide a commitment device to overcome the free rider problem associated with a public good like knowledge. Beyond this, the concentration of

impact evaluation funds in a small number of international institutions promotes rising standards of research excellence; can cluster studies in ways that accelerate learning; is highly efficient in terms of economics of scope and scale; and facilitates cross-sector and cross-disciplinary advances in methods and findings.

A Bright Future for Aid?

The future of aid is to build knowledge collectively. Foreign aid is better suited to funding pilot programs and impact evaluations than it is to financing domestic services or investments. The former supports institutional development and sustainable benefits; the latter risks undermining domestic political processes and stops sustaining services when funding ends.

An important part of aid's future is also to contribute toward and engage more in collective multilateral initiatives. The advantages in terms of driving rigorous standards of evidence; clustering studies; learning across contexts, sectors, and disciplines; and efficiencies in administration are manifest.

To confront the free rider problem directly, countries that provide foreign aid should pay a small levy on their disbursements that would be dedicated to impact evaluations – preferably with an independent entity like 3ie but also possibly with trust funds at multilateral agencies or in research centers around the world. If a significant number of countries would agree to make binding commitments to contribute 0.1% of their annual aid disbursements to 3ie, for example, they could assure stable funding for impact evaluation research, reduce the tendency of countries to be free riders, and accelerate progress in learning and accountability. Ideally, countries that use this policy evidence would also make commitments – say between $1 million and $50 million each year based on their domestic revenue capacity. This would be easier for many countries if 3ie were to obtain status as an international organization – with legal standing to receive funds from governments through official channels rather than as a simple non-profit organization.

In the last two decades, the world has seen exponential growth in the production of rigorous impact evaluations, increasing numbers of professionals capable of conducting evaluations, and significant improvements in methods. The growing number and quality of evaluations are making both technical contributions to program design and implementation, and helping governments fulfill a fundamental responsibility to use taxpayer money well. Over this same period, the world has seen a disruption in the conventional model of aid, and an active

search by aid agencies for their future role. As we have argued above, a particularly appropriate and essential future role is to provide predictable, meaningful financial support to institutions like the International Initiative for Impact Evaluation.

References

3ie (International Initiative for Impact Evaluation). 2014. Evidence, Influence, Impact: Annual Report 2013. New Delhi: International Initiative for Impact Evaluation.

Banerjee, A., Duflo, E., Imbert, C., and Pande. 2009-Present. "Empowering Female Leaders and Voters in Rajasthan, India." Cambridge, MA: Abdul Latiff Jameel Poverty Action Lab. http://www.povertyactionlab.org/evaluation/empowering-female-leaders-and-voters-rajasthan-india

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat A. L. B., Sturdy, J. and Vermeersch, C. M. J. 2011. "Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation." *The Lancet*, 377(9775): 1421-1428. http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(11)60177-3/abstract

Bass, J. K., Annan, J., Murray, S. M., Kaysen, D., Griffiths, S., Cetinoglu, T., Wachter, K., Murray, L. K., and Bolton, P. A. 2013. "Controlled Trial of Psychotherapy for Congolese Survivors of Sexual Violence."*New England Journal of Medicine* 368, no. 23 (2013): 2182-2191. http://www.nejm.org/doi/full/10.1056/NEJMoa1211853

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R. and Topalova, P. 2009."Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics*, 124(4): 1497-1540. November.

Bifulco, R. 2012. "Can Nonexperimental Estimates Replicate Random Assignment in Evaluations of School Choice? A Within-Study Comparison," *Journal of Policy Analysis and Management*. pp. 729-751. Summer.

Bosman, J. 2010. "City Will Stop Paying the Poor for Good Behavior." *New York Times*, March 30, 2010. http://www.nytimes.com/2010/03/31/nyregion/31cash.html?_r=0.

Cameron, D. and Mishra, A. 2014. "How Much Have We Learned? The Growth of Impact Evaluation for International Development." Mimeo. Washington DC: International Initiative for Impact Evaluation.

Cody, S. and Asher, A. 2014. "Smarter, Better, Faster: The Potential for Predictive Analytics and Rapid-Cycle Evaluation to Improve Program Development and Outcomes." Improving Safety Net and Work Support, the Hamilton Project. Washington, DC: The Brookings Institution. http://www.brookings.edu/~/media/research/files/papers/2014/06/19_hamilton_policies_addressing_poverty/predictive_analytics_rapid_cycle_evaluation_cody_asher.pdf.

Davis, B., Gaarder, M., Handa, S., and Yablonski, J. 2012. "Evaluating the impact of cash transfer programmes in sub-Saharan Africa: an introduction to the special issue." *Journal of Development Effectiveness* 4(1): 1-8. [Special Section].

De Allegri, M., Pokhrel, S., Becher, H., Dong, H., Mansmann, U., Kouyatè, B., Kynast-Wolf, G., Gbangou, A., Sanon, M., Briges, J., and Sauerborn, R. 2008. "Step-wedge Cluster-randomised Community-based Trials: An Application to the Study of the Impact of Community Health Insurance." *Health Research Policy and Systems* 6:10. http://www.health-policy-systems.com/content/6/1/10.

Desai, R. M. and Kharas, H. 2010. "The Determinants of Aid Volatility." Global Economy & Development, Working Paper 42. Washington, DC: The Brookings Institute.

http://www.brookings.edu/~/media/research/files/papers/2010/9/aid%20volatility%20desai%20kharas/09_aid_volatility.pdf

Development Channel Staff. 2012. "Question of the Week: Are Randomized Controlled Trials a Good Way to Evaluate Development Projects?" *Development Channel.* Washington, DC: Council of Foreign Relation. http://blogs.cfr.org/development-channel/2012/08/10/question-of-the-week-are-randomized-controlled-trials-a-good-way-to-evaluate-development-projects/

Evaluation Gap Working Group. 2006. "When Will We Ever Learn?" William D. Savedoff, Ruth Levine, and Nancy Birdsall, Co-Chairs, Center for Global Development, Washington, DC May 25. www.cgdev.org/section/initiatives/_active/evalgap

Gaarder. M. and Bartsch, U. 2014. "The Second Wave of Independence: Shopping for Solutions," Policy Research Working Paper 7069, Independent Evaluation Group. Washington, DC: World Bank.

Garbarino, S., and Holland, J. 2009. "Quantitative and Qualitative Methods in Impact Evaluation and the Measurement of Results," Governance and Social Development Resource Center. London: Overseas Development Institute. http://www.gsdrc.org/docs/open/EIRS4.pdf.

Gertler, P. and Giavagnoli, P. 2014. "Rewarding Provider Performance to Enable a Healthy Start to Life: Evidence from Argentina's Plan Nacer." World Bank Policy Research Working Paper No. 6884. Washington, DC: World Bank. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2439991.

Goldstein, M. and McKenzie, D. 2013. "The International Rescue Committee's approach to impact evaluation." Development Impact Blog. Washington, DC: World Bank. http://blogs.worldbank.org/impactevaluations/international-rescue-committees-approach-impact-evaluation

Gosden, T., Forland, F., Kristiansen, IS., Sutton, M., Leese, B., Giuffrida, A., Sergison, M., and Pedersen, L. 2001. "Impact of payment method on behaviour of primary care physicians: a systematic review." *Journal of Health Services Research & Policy* 6(1): 44-55.

Grindle, M. S. (2004). "Good enough governance: poverty reduction and reform in developing countries." *Governance*, 17(4): 525-548.

Guerrero, G., Leon, J., Zapata, M., Sugimaru, C., and Cueto, S. 2012. "What works to improve teacher attendance in developing countries? A systematic review." London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Habyarimana, J. and Jack, W. 2009. "Heckle and Chide: Results of a randomized road safety intervention in Kenya." Working Paper Number 169. Washington, D.C: Center for Global Development

Haskins, R. and Margolis, G. 2014. *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy.* Washington, DC: Brookings Institution.

Hatanga, P. and Herbert, T. 2011. "Uganda Tests Ability of PES to Deliver Benefits for Biodiversity." in *Creating New Values for Africa: Emerging Ecosystem Service Markets.* Washington, DC: Ecosystem Marketplace. pp. 54-57.

http://www.undp.org/content/dam/undp/library/Environment%20and%20Energy/biodiversity/PES%20in%20Africa_12-29-11.pdf

Haushofer, J. and Shapiro, J. 2013. "Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya." http://www.princeton.edu/~joha/publications/Haushofer_Shapiro_UCT_2013.pdf.

Independent Evaluation Group (IEG). 2012. "World Bank Group Impact Evaluation: Relevance and Effectiveness." Washington, DC: World Bank. http://ieg.worldbankgroup.org/Data/reports/impact_eval_report.pdf

International Initiative for Impact Evaluation (3ie). "Systematic Reviews." Last modified 2012. http://www.3ieimpact.org/en/evidence/systematic-reviews/?q=&title=&author=&published_from=&published_to=&funded=on

Jacquet, P. 2006. "Evaluations and Aid Effectiveness." in N. Birdsall, ed. *Rescuing the World Bank: a CGD working group report and selected essays*. Washington, DC: Center for Global Development. pp. 171-182.

Karlan, D. and Goldberg, N. 2007. "Impact Evaluation for Microfinance: Review of Methodological Issues." Washington, DC: World Bank. http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_07.pdf.

Kavanagh, Matthew M. 2014. "The Politics of Transition and the Economics of HIV: AIDS & PEPFAR in South Africa." Health Global Access Project, University of Pennsylvania. January. http://www.health-e.org.za/wp-content/uploads/2014/01/Politics-of-Transition-Report-AIDS-South-Africa.pdf

Kharas, H. 2008. "Measuring the Cost of Aid Volatility" Wolfensohn Center for Development, Working Paper 3. Washington, DC: The Brookings Institute. http://www.brookings.edu/~/media/research/files/papers/2008/7/aid%20volatility%20kharas/07_aid_volatility_kharas.pdf

Kiplesund, S. and Morton, M. 2014. "Gender-based violence prevention: lessons from World Bank impact evaluations." enGender Impact: the World Bank's Gender Impact Evaluation Database. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/2014/04/19456525/gender-based-violence-prevention-lessons-world-bank-impact-evaluations

LSHTM (London School of Hygiene & Tropical Medicine). 1979-Present. *Karonga Prevention Study*. London. http://www.lshtm.ac.uk/eph/ide/research/kps/#

MCC (Millennium Challenge Corporation). 2013. "Independent Evaluations." http://www.mcc.gov/pages/results/evaluations (Last modified 2013).

MCC (Millennium Challenge Corporation). n.d. "The First Five Evaluations." Washington, DC: Millennium Challenge Corporation. http://www.mcc.gov/pages/features/first-five-evaluations

Mejía, F. 2014. "Laptops, children and Darth Vader" *Development that works. Effectiveness Blog.* Washington, DC: Inter-American Development Bank. http://blogs.iadb.org/desarrolloefectivo_en/2014/09/19/laptops-children-darth-vader/

Office of Adolescent Health. "Teen Pregnancy Prevention." Last modified 2014. Washington, DC: U.S. Department of Health & Human Services. http://www.hhs.gov/ash/oah/oah-initiatives/tpp/

Pritchett, L. 2002. "It pays to be ignorant: a simple political economy of rigorous program evaluation." *The Journal of Policy Reform,* 5(4): 251-269.

Pritchett, L. 2013. "RCTs in Development, Lessons from the Hype Cycle." *Global Development: Views from the Center*. Washington, DC: Center for Global Development. http://www.cgdev.org/blog/rcts-development-lessons-hype-cycle

Result for Development. 2013. "RFD and Harvard Kennedy School Awarded $8.1 Million to Research Government Accountability and Transparency in Indonesia and Tanzania." *Press Release*. Washington, DC: Results for Development. http://r4d.org/about-us/press-room/r4d-and-harvard-kennedy-school-awarded-81-million-research-government-accountabi

Rossi, Peter Henry, Mark W. Lipsey, and Howard E. Freeman. 2003. *Evaluation: A systematic approach*. Sage publications.

Savedoff, W. D. 2014. "End the Evaluation Wars: A Plea to Shift from the Abstract to the Specific." *Global Development: Views from the Center*. Washington, DC: Center for Global Development. http://www.cgdev.org/blog/end-evaluation-wars-plea-shift-abstract-specific

Schochet, P., Puma, M., and Deke, J. 2014. "Understanding Variation in Treatment Effects in Education Impact Evaluations: An Overview of Quantitative Methods," National Center for Education Evaluation and Regional Assistance. http://ies.ed.gov/ncee/pubs/20144017/pdf/20144017.pdf.

Strange, A., Parks, B. C., Tierney, M. J., Fuchs, A., Dreher, A., Ramachandran, V. 2013. "China's Development Finance to Africa: A Media-Based Approach to Data Collection." CGD Working Paper 323. Washington, DC: Center for Global Development. http://www.cgdev.org/publication/chinas-development-finance

Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., Rosenthal, M. B., and Sermeus, W. 2010. "Systematic review: effects, design choices, and context of pay-for-performance in health care." *BMC Health Services Research* 10(1):247. http://www.biomedcentral.com/1472-6963/10/247

Villamor, M. 2014. "Experiments in development: Time to rethink RCTs?" *Devex*. Washington, DC: DEVEX. https://www.devex.com/news/experiments-in-development-time-to-rethink-rcts-84344

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G. Vojtkova, M., Davies. P., Bhavsar, A., Eyers, J., Koehlmoos, T. P., Perricrew, M., Valentine, J. C. and Tugwell, P.2012. "How to do a good systematic review of effects in international development: a tool kit," *Journal of Development Effectiveness*, 4(3):359-387, September.

White, H. 2009. "Theory-based impact evaluation: principles and practice." *Journal of development effectiveness,* 1(3): 271-284. http://www.tandfonline.com/doi/abs/10.1080/19439340903114628#.VElwIvl4pNM

White, H. 2013. "An introduction to the use of randomised control trials to evaluate development interventions," *Journal of Development Effectiveness,* 5(1):30-49.

http://www.tandfonline.com/doi/full/10.1080/19439342.2013.764652#.VElwq_l4pNN

White, H., & Waddington, H. 2012. "Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews." *Journal of Development Effectiveness*, 4(3): 351-358.

World Bank. "Data & Research." Last modified 2014. Washington, DC: World Bank. http://go.worldbank.org/69LQQAL850

World Bank. 2015. *World Development Report 2015: Mind, Society, and Behavior*. Washington, DC: World Bank.

World Bank. 2004. *World Development Report 2004: Making Social Services Work*. Washington, DC: World Bank.

Zamboni, Y. and Litschig, S. 2014. "Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil." Working Papers 554, Barcelona: Barcelona Graduate School of Economics. https://ideas.repec.org/p/bge/wpaper/445.html.