Aid at the frontier: building knowledge collectively

Ruth Levine\textsuperscript{a} and William Savedoff\textsuperscript{b}\textsuperscript{*}

\textsuperscript{a}Global Development and Population Program, William and Flora Hewlett Foundation, 2121 Sand Hill Road, Menlo Park, CA 94025, USA; \textsuperscript{b}Center for Global Development, 2055 L Street, NW, Washington, DC 20036, USA

This paper articulates how programme evaluation generally, and impact evaluation specifically, contributes to good governance – not by replacing politics, but by informing it. We argue that institutions with the mandate to accelerate progress in the developing world through aid transfers are particularly well suited to fund impact evaluations. We argue, in fact, that funding impact evaluations through a collective vehicle like the International Initiative for Impact Evaluation (3ie) should be a primary focus of foreign aid. Finally, we highlight the conditions needed and the additional efforts required to yield the full benefits of collective investment in finding out what works.

Keywords: evaluation; impact evaluation; public goods; public policy; aid effectiveness; development

Introduction

All governments face the same problem: how can they know when they try to help citizens if they are successful or are, instead, wasting tax money and slowing social and economic progress? Obtaining that knowledge is hard and even considered a quixotic ambition in the data-poor environments of many middle- and low-income countries. Taking time to learn how well government programmes work has also been criticised as a technocratic sideshow to the main stage of politics.

Nevertheless, throughout the world policymakers and citizens are recognising that the very legitimacy of public sector institutions is jeopardised by their inability to demonstrate their impact and, when necessary, to change course. 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 more exigent. Evaluation is increasingly seen as an activity that is intrinsic to good government.

While the evaluation of public policies and programmes relies on innovations and experiences developed over more than half a century (Rossi, Lipsey, and Freeman 2003, 9–20), in recent years, researchers and practitioners have greatly expanded the range of methods applied to programme evaluation in low- and middle-income countries. Building on experiences in industrialised countries, international champions of impact evaluation like Howard White and his colleagues at the International Initiative for Impact Evaluation (3ie), along with the broader community of academics, officials, staff and implementers, have pursued innovative evaluation approaches that are appropriate for varied contexts in middle- and low-income countries. Pioneering figures in a few countries like South Africa, Mexico, Colombia, Brazil, Indonesia, Rwanda and Kenya have committed to

\textsuperscript{*}Corresponding author. Email: wsavedoff@cgdev.org
evaluation as an instrument of accountability to voters and a means of fulfilling their executive responsibilities.

In this paper, we seek to articulate how programme evaluation generally, and impact evaluation specifically, contributes 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 are particularly well suited to fund impact evaluations. We argue, in fact, that funding policy-relevant impact evaluations through a collective entity like the 3ie should be one of their primary activities. Finally, we highlight the conditions needed and the additional efforts required to yield the full benefits of collective investment in finding out what works. Throughout, in honour of Howard White’s leadership in the field, we reference some of his many contributions to the literature and the practice of impact evaluation.

Politics rules, but evaluation still holds much promise

No matter how much some people would like ‘value-neutral’ decision tools to guide public policy, government priorities are (rightly) established through the wonderful and messy human process referred to as ‘politics.’ Evidence, knowledge and technical expertise are part of this process, but do not (and should not) determine the policy outcome. Rather, evidence is itself contested in this arena and enters and influences the political process alongside other social phenomena like interests, power and prejudice. Yet evidence still has a role to play. It influences debate at opportune moments; it can draw attention and marshal responses to serious problems; and within public systems, it can influence managerial choices, regulatory decisions and policy design.

Many kinds of analysis are useful to improving public policy. Rigorous factual analysis can detect how seemingly well-designed programmes lose their way in basic implementation (White 2009). Different kinds of evaluations can provide valuable information about the cost of interventions, targeting strategies and system outputs. They can shed light on institutional strengths and weaknesses that influence the sustainability of any effort. Other kinds of evaluations can reveal the meaning and interpretation of change as experienced by beneficiaries themselves. Impact evaluations can show whether programmes that were implemented improved outcomes, provided gains large enough to be worth the cost and were more effective than alternative uses of the same resources.

Impact evaluation is an evolving and growing field

In 2006, the Center for Global Development (CGD) Evaluation Gap Working Group pointed out in its report, ‘When Will We Ever Learn: Improving Lives through Impact Evaluation,’ that impact evaluations of both government and donor-funded programmes were systematically underfunded despite their potential value (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 programmes 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. Other reasons advanced 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).
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.

Impact evaluation 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. Researchers have developed a range of techniques to address this problem, including statistical methods like instrumental variables, difference-in-differences, matching and regression discontinuity. Qualitative studies which gather data from non-participating individuals or organisations are similarly engaged in extracting information that is used to account for confounding factors by constructing a plausible counterfactual (Rossi, Lipsey, and Freeman 2003; White 2009).

Studies in the last decades have pioneered novel approaches to estimating impact. Development economists have adapted randomised control trial methods to compare health, educational, behavioural 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). Researchers have also addressed the quality or intensity of programme implementation in evaluation design (Schellenberg, Bobrova, and Avan 2012); 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, Puma, and Deke 2014). White has challenged researchers to insist on clarity over causal chains and theory when conducting impact evaluations so as to understand not only what works but also why (White 2009). While the methodological challenges are far from being solved, impact evaluation has proven to be a field that innovates persistently and adopts new solutions quickly.

Since 2006, the impact evaluation community has been bold as well in its choice of subject matter. Early applications of impact evaluation focused primarily on health and educational outcomes that had reasonably standard definitions and could be measured and compared at the individual level. Today, impact evaluators regularly tackle a broader array of topics, including: gender-based violence (Kiplesund and Morton 2014), road traffic accidents (Habyarimana and Jack 2011), forest preservation (Hatanga and Herbert 2011), corruption (Zamboni and Litschig 2014), trauma following rape (Bass et al. 2013), savings behaviour (Karlan and Goldberg 2007), income gains (Hausbofer and Shapiro 2013), women’s empowerment (Beaman et al. 2009), teacher absenteeism, health worker performance and health outcomes (Basinga et al. 2011) and accountability through citizen action (Results for Development 2013).

Beyond methods and scope, the impact evaluation community has simply been very busy doing impact evaluations. Researchers at 3ie have compiled the most complete database of published impact evaluations available and found 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, Mishra, and Brown 2015). The Abdul Jameel Latif Poverty Action Lab (JPAL) and Innovations for Poverty Action (IPA) are relatively new research centres dedicated to conducting impact evaluations, and in 2014 they listed 567 and 220 ongoing studies, respectively. Other research centres 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.

Many agencies are involved in funding or commissioning this work, although the sources are highly concentrated and weakly coordinated. Since 3ie began in 2009, it 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. Organisations 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 scale of programmes in developing countries which are 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

Figure 1. Impact evaluations published per year, 1981–2012.
Source: Reproduced from Cameron, Mishra, and Brown (2015).
dedicated government units concerned with evaluating public programmes and commis-
sioning 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.

The findings from impact evaluations produce concentric circles of benefit. First and
foremost, the findings can be used by those implementing a programme. Depending on
the findings, implementers can continue more confidently with successful approaches,
proceed to modify less successful ones and redirect their energy towards new strategies
when faced with failures. A second circle of benefits accrues to those working with
similar problems and populations in other places. This next ring of people can build
programmes around interventions that have shown success or at least avoid mistakes made
by others.

Impact evaluations generate a third ring of concentric benefits by expanding the body
of evidence that helps steer governments and funders towards more effective approaches
and away from less effective ones. Part of this ring consists of systematic reviews of
impact evaluations which provide an overview of the evidence across different contexts,
isights regarding which theories are more useful and opportunities to assess generalisa-
brility (White and Waddington 2012; Waddington et al. 2012). 3ie itself has produced 15
systematic reviews on topics ranging from the impact of field schools on agriculture
productivity to interventions to prevent HIV through behaviour change.2 The Interna-
tional Development Coordinating Group of the Campbell Collaboration began to publish sys-
tematic reviews in 2012 and now has 40 studies on its website. The outermost ring of the
circle of benefits constitutes the general knowledge of how programmes and people
behave, useful frames for thinking about problems and increased capacities to think,
research, adapt, interpret and assess.

Future progress in impact evaluation is threatened

Despite its recent growth, future investments in impact evaluation are not guaranteed. To
garner support, impact evaluations need to be credible by maintaining high standards of
analytical rigor, useful by investigating relevant policy questions and efficient in terms of
costs and time. Impact evaluation may be at the top of a ‘hype cycle’ (Pritchett 2013) if it
cannot deliver on its promised benefits. 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. Methodological debates among professional evaluators and researchers,
particularly over randomisation, may also inadvertently strengthen those who deride
research as sterile and unproductive.3

Tensions between the interests of researchers and the goals of policymakers may also
temper enthusiasm for impact evaluations. Researchers have incentives to carry out
studies that can be published in prestigious journals and enhance their standing among
their peers. Researchers may not always find the questions being asked by policymakers
to be interesting or researchable. Furthermore, studies that replicate existing studies to test
their robustness and generalisability are of great importance to advising policymakers but
of less value in academia. It is to the credit of the Journal of Development Effectiveness,
under the guidance of Howard White, that it adopted a policy to explicitly solicit and
promote the publication of replication studies.
Finally, relatively few public or private funders have invested in impact evaluation. Although significantly more funding is available today than before, 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 few funders. Unless a broader range of governments and organisations can be convinced of the need for collectively providing long-term stable funding to this knowledge-building endeavour, enormous opportunities will be lost.

Various solutions may be found to address these threats and problems, but fundamental to almost all of them is obtaining broad financial support, particularly through a collective, fit-for-purpose mechanism like 3ie. That is how foreign aid can advance evaluation as an essential element of good governance.

**Aid is uniquely suited to impact evaluations**

Foreign aid can be helpful in many ways, but it is particularly useful for learning how to make public programmes more effective. In this case, we are referring to public programmes financed 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 well 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 (Hjertholm and White 2000; White and Dijkstra 2013). This is no longer the case. Over the past 20 years, most low- and middle-income countries have grown economically. They are attractive to private investors and their governments have experienced substantial growth in their own revenues. Overseas development assistance has fallen as a share of donor country gross national income (GNI) from 0.5 per cent in the 1960s to about 0.3 per cent 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. 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 per cent of GNI in low-income countries (a shrinking category of countries) and represent about 3 per cent of GNI in the middle-income countries that are still receiving aid (See Figure 3).

The declining importance of aid may be a good thing because aid can be volatile and disruptive (Desai and Kharas 2010; Kharas 2008). Nevertheless, aid still has an important role to play. Despite progress in many places, problems of poor health, poverty and slow growth persist in many parts of the world and we have new challenges, such as climate disruption and the rise of non-communicable diseases. Increasingly, we also recognise that sustainable development requires changes in governance, social norms and incentives (Grindle 2004; World Bank 2004, 2015).
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 GNI for countries within each income group. Countries were included 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.
Recognising 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 specialise 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. The third path is to contribute with knowledge. While the third path might be dismissed as ‘merely doing research,’ knowledge from such research has the potential to be one of the greatest sources 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 $1000 per capita does not need international agencies to fund its health system. Yet Argentina sought a loan for extending provincial health insurance in order to benefit from the World Bank’s technical assistance and support to do a rigorous impact evaluation of the programme (Gertler and Giavagnoli 2014). By generating knowledge with relatively few dollars, the project improved the effectiveness of Argentina’s domestic health spending far beyond the amount of aid or loans that any organisations could have directly offered or sustained.

The main advantage of foreign aid is its ability to mobilise 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 10 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, mobilise and engage the right technical experts. Since its founding, 3ie 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.

Furthermore, the goals for foreign assistance related to improving domestic governance can be better achieved through financing impact evaluations than through direct projects. Foreign funding can be in tension with democratic processes when it distorts domestic priorities. By contrast, aid that evaluates public programmes provides information to both governments and citizens that instead reinforces accountability and strengthens governance.

We recognise 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 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 assisting a small group of
extremely low-income countries, 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 programmes – 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 programmes – than financing investment and services – that alter and substitute for domestic funding. Foreign aid can mobilise 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 that it fosters the development of domestic institutions and capacity to research, assess and learn about public programmes, it can have even longer lasting effects.

**Collective is more effective**

Though foreign investment in impact evaluation is helpful, collective action to fund impact evaluations is even more likely to succeed at advancing knowledge and improving public programmes. 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. Second, 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 endeavour.

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. One government 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.

This public good characteristic of knowledge, however, also means that if each individual, organisation 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. Aid programmes are subject to this same limitation. They regularly face pressures to generate information that is useful to their programmes and today’s policy questions. In so doing, they forego opportunities to invest in studying programmes 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. One 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 programmes – similar to international commitments to contribute 0.7 per cent 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 tendency to underinvest in public goods was a key argument for proposing the creation of 3ie. Yet funding for 3ie in its first 6 years remains concentrated, showing 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 2014).

The collective decision to adequately fund impact evaluations may be hard, but creating an institution to receive and apply those funds has many benefits. By working internationally through an institution like 3ie, research grant decisions are less easily influenced by constituencies within particular countries and can draw from a larger community of disinterested parties when selecting topics or awarding grants. Furthermore, grants can be clustered around related questions to improve the reliability and generalisability of findings. Having a number of studies with similar findings gives policymakers and researchers greater confidence in the findings.

As a collective institution, 3ie can also be more efficient because of scale economies in administering a grant programme. Most agencies are too small to have staff with the expertise and time available for 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. Only grant programmes with a significant number of applications can afford to convene enough people with sufficiently specialised knowledge to assess proposals.

Another advantage for a collective institution like 3ie is that it can undertake grant review in ways that foster 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.

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 generalisable and that countries may be less likely to absorb lessons from studies financed, let alone conducted, by other countries. Collective financing of impact evaluations that are commissioned through an independent organisation runs the risk of losing opportunities to improve projects when evaluators can engage with project designers and implementers (Jacquet 2006; Gaarder and Bartsch 2014). The relevance of impact evaluations is also a regular concern, one which requires the institutions commissioning impact evaluations 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 address these concerns. We already have a number of ways of addressing generalisability and relevance. These include the following:
• mechanisms to promote ‘practice-relevant’ impact evaluations, such as 3ie’s ‘Policy Window’;
• clustering studies around common questions so that evidence from different contexts can be used to assess external validity (as occurs in 3ie’s ‘Thematic Window’);
• programmes 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;
• 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;
• 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 their substantive engagement; and
• 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 towards 3ie or other collective 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 programmes 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 towards 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 to 3ie but also possibly with trust funds at multilateral agencies or in regional research. If a significant number of countries would agree to make binding commitments to contribute 0.1 per cent 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, the low- and middle-income countries that use this policy evidence would also make modest 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 organisation – with legal standing to
receive funds from governments through official channels rather than as a simple non-profit organisation.

In the last two decades, the number of rigorous impact evaluations has grown, the number of professionals capable of conducting evaluations has risen and methods of research have improved. This has contributed to improvements in programmes and helped many governments address their responsibilities to use taxpayer money well. Over this same period, the conventional model of aid has been experiencing disruption, prompting 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 impact evaluation. The appropriate instrument for this support is the 3ie and, after Howard White’s term as Executive Director, it has a proven track record that justifies investing in it with confidence.

Acknowledgements
The authors would like to thank Albert Alwang for his research assistance; participants in the conference ‘Improving lives through better evidence: an acknowledgement of the contribution of Howard White to international development,’ at which Ruth Levine presented this paper in October 2014, for their feedback; and anonymous reviewers who commented on an earlier CGD policy paper version and on the draft submitted to this journal.

Disclosure statement
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 any funding from 3ie or been involved in its management since 2010. Ruth Levine was appointed Institutional Representative for Foundations and NGOs to 3ie’s Board of Commissioners in April 2015. The Hewlett Foundation is one of several founding funders of 3ie.

Notes
1. In this paper, ‘foreign aid’ will refer to financing provided by members of the OECD/Development Assistance Committee and multilateral development banks to low- and middle-income countries. Some of the issues raised apply also to large-scale private philanthropy.
2. 3ie website, accessed July 8, 2015. http://www.3ieimpact.org/en/evidence/systematicreviews/?funded=True&q=&sort_by=alphabet&page=3.
3. Examples of the debate include Development Channel Staff (2012), Villamor (2014) and Savedoff (2014).
4. 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.
5. This feature is called ‘nonrivalry in consumption’ in the economics literature. A second feature – non-excludability – is also required to characterise a pure public good.
6. One of the most dramatic examples of how knowledge has affected human well-being is the rise in human life expectancy (Kenny 2011).

References
3ie (International Initiative for Impact Evaluation). 2014. Evidence, Influence, Impact: Annual Report 2013. New Delhi: International Initiative for Impact Evaluation.
Basinga, P., P. J. Gertler, A. Binagwaho, A. L. Soucat, J. Sturdy, and C. M. Vermeersch. 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. doi:10.1016/S0140-6736(11)60177-3.
Bass, J. K., J. Annan, S. M. Murray, D. Kaysen, S. Griffiths, T. Cetinoglu, K. Wachter, L. K. Murray, and P. A. Bolton. 2013. “Controlled Trial of Psychotherapy for Congolese Survivors of Sexual Violence.” New England Journal of Medicine 368 (23): 2182–2191. doi:10.1056/NEJMoA1211853.

Beam, L. R., Chattopadhyay, E. Duflo, R. Pande, and P. Topalova. 2009. “Powerful Women: Does Exposure Reduce Bias?” Quarterly Journal of Economics 124 (4): 1497–1540. doi:10.1162/qjec.2009.124.4.1497.

Bifulco, R. 2012. “Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison.” Journal of Policy Analysis and Management 31 (3): 729–751. doi:10.1002/pam.20637.

Cameron, D. B., A. Mishra, and A. N. Brown. 2015. “The Growth of Impact Evaluation for International Development: How Much Have We Learned?” Journal of Development Effectiveness 1–21. doi:10.1080/19439342.2015.1034156.

Cody, S., and A. Asher. 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

Desai, R. M., and H. Kharas. 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? Improving Lives through Impact Evaluation. Co-chaired by W. D. Savedoff, R. Levine, and N. Birdsell. Washington, DC: Center for Global Development. http://cgdev.org.488elwb02.blackmesh.com/sites/default/files/7973_ file_WillWeEverLearn.pdf

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

Gertler, P. J., and P. I. Giavagnoli. 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

Grindle, M. S. 2004. “Good Enough Governance: Poverty Reduction and Reform in Developing Countries.” Governance 17 (4): 525–548. doi:10.1111/gove.2004.17.issue-4.

Habyarimana, J., and W. Jack. 2011. “Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya.” Journal of Public Economics 95 (11–12): 1438–1446. doi:10.1016/j.jpubeco.2011.06.008.

Hatanga, P., and T. Herbert. 2011. “Uganda Tests Ability of PES to Deliver Benefits for Biodiversity.” In Creating New Values for Africa: Emerging Ecosystem Service Markets, 54–57. Washington, DC: Forest Trends, The Katoomba Group and Ecosystem Marketplace. http://www.undp.org/content/dam/undp/library/Environment%20and%20Energy/biodiversity/PES%20in%20Africa_12-29-11.pdf

Haushofer, J., and J. Shapiro. 2013. Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya. Mimeo. Princeton: Princeton University. http://www.princeton.edu/~joha/publications/haushofer_shapiro_UCT_2013.pdf

Hjertholm, P., and H. White. 2000. “Foreign Aid in Historical Perspective.” Chap. 3 in Foreign Aid and Development: Lessons Learnt and Directions for the Future, edited by F. Tarp, 59–77. London: Routledge.

IEG (Independent Evaluation Group). 2012. World Bank Group Impact Evaluation: Relevance and Effectiveness. Washington, DC: World Bank.

Jacquet, P. 2006. “Evaluations and Aid Effectiveness.” In Rescuing the World Bank: A CGD Working Group Report and Selected Essays, edited by N. Birdsall, 171–182. Washington, DC: Center for Global Development.
Karlan, D., and N. Goldberg. 2007. *Impact Evaluation for Microfinance: Review of Methodological Issues*. Washington, DC: World Bank.  [http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752420884/Doing_ie_series_07.pdf](http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752420884/Doing_ie_series_07.pdf)

Kenny, C. 2011. *Getting Better: Why Global Development is Succeeding – and How We Can Improve the World Even More*. New York: Basic Books.

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](http://www.brookings.edu/~media/research/files/papers/2008/7/aid%20volatility%20kharas/07_aid_volatility_kharas.pdf)

Kiplesund, S., and M. Morton. 2014. “Gender-Based Violence Prevention: Lessons from World Bank Impact Evaluations.” In *Gender 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](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. Accessed July 8, 2015.  [http://www.lshtm.ac.uk/eph/ide/research/kps/#](http://www.lshtm.ac.uk/eph/ide/research/kps/#)

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. doi: 10.1080/1384128032000096832.

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](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](http://r4d.org/about-us/press-room/r4d-and-harvard-kennedy-school-awarded-81-million-research-government-accountabi)

Rossi, P. H., M. W. Lipsey, and H. E. Freeman. 2003. *Evaluation: A Systematic Approach*. London: 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](http://www.cgdev.org/blog/end-evaluation-wars-plea-shift-abstract-specific)

Schellenberg, J. A., N. Bobrova, and B. I. Avan. 2012. *Measuring Implementation Strength: Literature Review Draft Report*. London: London School of Hygiene and Tropical Medicine.  [http://researchonline.lshtm.ac.uk/1126637/](http://researchonline.lshtm.ac.uk/1126637/)

Schochet, P. Z., M. Puma, and J. Deke. 2014. *Understanding Variation in Treatment Effects in Education Impact Evaluations: An Overview of Quantitative Methods*. (NCEE 2014–4017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development.  [http://ies.ed.gov/ncee/edlabs](http://ies.ed.gov/ncee/edlabs)

Strange, A., B. C. Parks, M. J. Tierney, A. Fuchs, A. Dreher, and V. Ramachandran. 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](http://www.cgdev.org/publication/chinas-development-finance)

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](https://www.devex.com/news/experiments-in-development-time-to-rethink-rcts-84344)

Waddington, H., H. White, B. Snisltveit, J. G. Hombrados, M. Vojtkova, P. Davies, A. Bhavsar, et al. 2012. “How to Do A Good Systematic Review of Effects in International Development: A Tool Kit.” *Journal of Development Effectiveness* 4 (3): 359–387. doi: 10.1080/19439342.2012.711765.

White, H. 2009. “Theory-Based Impact Evaluation: Principles and Practice.” *Journal of Development Effectiveness* 1 (3): 271–284. doi: 10.1080/19439340903114628.

White, H. 2013. “An Introduction to the Use of Randomised Control Trials to Evaluate Development Interventions.” *Journal of Development Effectiveness* 5 (1): 30–49. doi: 10.1080/19439342.2013.764652.

White, H., and H. Waddington. 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. doi: 10.1080/19439342.2012.711343.

White, H. N., and A. G. Dijkstra. 2013. *Programme Aid and Development: Beyond Conditionality*. London: Routledge.
World Bank. 2004. *World Development Report 2004: Making Social Services Work*. Washington, DC: World Bank.

World Bank. 2014. *Data & Research*. Last modified 2014. Washington, DC: World Bank. [http://go.worldbank.org/69LQQAL850](http://go.worldbank.org/69LQQAL850)

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

Zamboni, Y., and S. Litschig. 2014. “Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil.” *Working Papers 554*. Barcelona: Barcelona Graduate School of Economics. [http://repositori.upf.edu/handle/10230/19859](http://repositori.upf.edu/handle/10230/19859)