

The New Transparency in Development Economics: Lessons from the Millennium Villages Controversy

Michael Clemens and Gabriel Demombynes

Abstract

The Millennium Villages Project is a high profile, multi-country development project that has aimed to serve as a model for ending rural poverty in sub-Saharan Africa. The project became the subject of controversy when weaknesses in its claims of success were exposed. The lively ensuing debate offers lessons on three recent mini-revolutions that have swept the field of development economics: the rising standards of evidence for measuring impact, the “open data” movement, and the growing role of the blogosphere in research debates.

The definitive version of this paper is forthcoming in *The World Economy*.

JEL Codes: F35, O12, O22

Keywords: Millennium Villages Project, impact evaluation, transparency, open data.

The New Transparency in Development Economics: Lessons from the Millennium Villages Controversy

Michael Clemens
Center for Global Development

Gabriel Demombynes
World Bank and NYU Financial Access Initiative

We thank Nabil Hashmi for research assistance, and we appreciate comments and reactions from Søren Davidsen, Justin Sandefur, and Eric Verhoogen. This text represents the views of the authors alone and does not necessarily represent the views of the Center for Global Development or the World Bank, their boards, or their funders.

CGD is grateful for support of this work from its funders and board of directors.

Michael Clemens and Gabriel Demombynes . 2013. "The New Transparency in Impact Evaluation: Lessons from the Millennium Villages Controversy." CGD Working Paper 342. Washington, DC: Center for Global Development.
<http://www.cgdev.org/publication/new-transparency-impact-evaluation-lessons-millennium-villages-controversy>

Center for Global Development
1800 Massachusetts Ave., NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

The Center for Global Development is an independent, nonprofit policy research organization dedicated to reducing global poverty and inequality and to making globalization work for the poor. Use and dissemination of this Working Paper is encouraged; however, reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License.

The views expressed in CGD Working Papers are those of the authors and should not be attributed to the board of directors or funders of the Center for Global Development.

Introduction

Three minor revolutions are reshaping the field of development economics. The first concerns standards of evidence. Journal editors and project funders increasingly stress rigorous impact evaluation (IE). The second revolution is in the medium of debate. Discourse that traditionally happened only within journals and conferences now spills over into the more rapid and democratic medium of blogs. The third revolution is in the ease and cost of accessing research data. We argue that these upheavals deserve applause, because they can bring unprecedented degrees of transparency to discussions on the impact of development projects.

This new transparency matters because the way thinking in development economics is shaped affects the choices made by policymakers around the world. Despite widespread progress, more than one billion people still survive on less than \$1.25 a day. The World Bank has advocated a global target of ending such extreme poverty by 2030, and fierce debates rage about the right choices of programs and policies to approach that goal.

As a case study we highlight a public debate in which we participated regarding the Millennium Villages Project (MVP). The Project is a large experimental intervention at 15 sites across rural sub-Saharan Africa which is designed to show that “people in the poorest regions of rural Africa can lift themselves out of extreme poverty in five years’ time” (MVP 2007). We focus on the MVP because of its prominence and because the project illustrates all three elements of the new transparency trend.

We begin with a broad explanation of impact evaluation. Second, we describe the MVP and its initial evaluation design. Next, we describe our critique of the MVP evaluation and provide a narrative of how a debate about the MVP evaluation unfolded across multiple forms of media, with blogs and data access profoundly shaping the discussion. Finally, we consider five lessons from the MVP controversy.

1. The impact evaluation problem

It is helpful to start with a statement of the evaluation problem. What does it mean to say that a program has an “effect” or “impact”? Let us say we are evaluating a job training program and trying to assess its impact on the income of individual i . We can imagine two possible scenarios. In one scenario, individual i participates in the training program, and her income changes by some value, which we will call $Y_i(1)$. In another scenario, individual i does not participate in the program and has income which changes by a different value, $Y_i(0)$. We define the effect of the program on individual i 's income as $Y_i(1) - Y_i(0)$. In words, the effect of the program on income is simply the difference between the change in income the individual experiences if she is in the program and the change in income she experiences if she is not in the program.

The challenge is that we do not observe both $Y_i(1)$ and $Y_i(0)$. Either the individual is in the program or she is not. If she is a program participant, we can observe $Y_i(1)$ by asking her directly or obtaining some other record of her income before and after the training. But then $Y_i(0)$ will exist as only a hypothetical construct. This what-would-have-happened scenario is termed the “counterfactual.”

In the more general problem, Y could be the change in any outcome of interest—such as a test score, a malnutrition indicator, or a measure of happiness—and the unit of analysis could be the individual or something else, such as the household or village. The challenge in evaluation is to develop an estimate for $Y_i(1) - Y_i(0)$ without observing both values. Any serious attempt at impact evaluation requires some explicit way to estimate what would have happened if the intervention had never happened. There are a number of different approaches, all of which rely on finding a stand-in for the counterfactual. These include differences-in-differences, regression discontinuity, randomized controlled trials (RCTs), and propensity-score and other forms of matching.¹

Across economics, IE has become more common, and the use of RCTs in particular has become widespread in development in recent years. The RCT-based evaluation of the PROGRESA program in Mexico in the late 1990s received great attention, along with several RCT studies by Michael Kremer and associates at Harvard and Esther Duflo and associates at MIT. Subsequently, many scholars across academia and institutions like the World Bank have focused their work on rigorous IE.

2. The Millennium Villages evaluation

In this context, the Millennium Villages Project was launched, first in Sauri, Kenya in 2004 and then later at a number of sites across sub-Saharan Africa. The MVP is an experimental intervention to promote economic development in clusters of very poor rural villages across Africa. It is a joint project of the United Nations Development Program, Columbia University’s Earth Institute, and Millennium Promise, a non-governmental organization.

a. The intervention

The MVP deploys a package of interventions, including distribution of fertilizer and insecticide-treated bednets, school construction, HIV testing, microfinance, electric lines, road construction, and water and irrigation. The mix of interventions differs in each village cluster. The project was designed to break villages out of poverty traps (Sachs et al. 2004)

¹ In practice, we typically estimate effects for groups rather than individual units. We recommend *Impact Evaluation in Practice* by Paul Gertler et al. (2011), available for free download on the World Bank website, as a useful introduction to IE techniques.

and “make the investments in human capital and infrastructure required to achieve self-sustaining economic growth” (MVP 2007).

The MVP is broadly similar to the “integrated rural development” (IRD) approach that was implemented in many countries in the 1970s and early 1980s before a consensus developed that it did not work. As one expert wrote, “Billions and billions of dollars went into hundreds of such programs without hardly any sustainable impact or poverty or institutions anywhere” (Binswanger-Mkhize 2011). The troubled history of IRD invites skepticism that the approach can spark sustained economic development, making it particularly important that projects of this type undergo rigorous IE.²

b. The evaluation

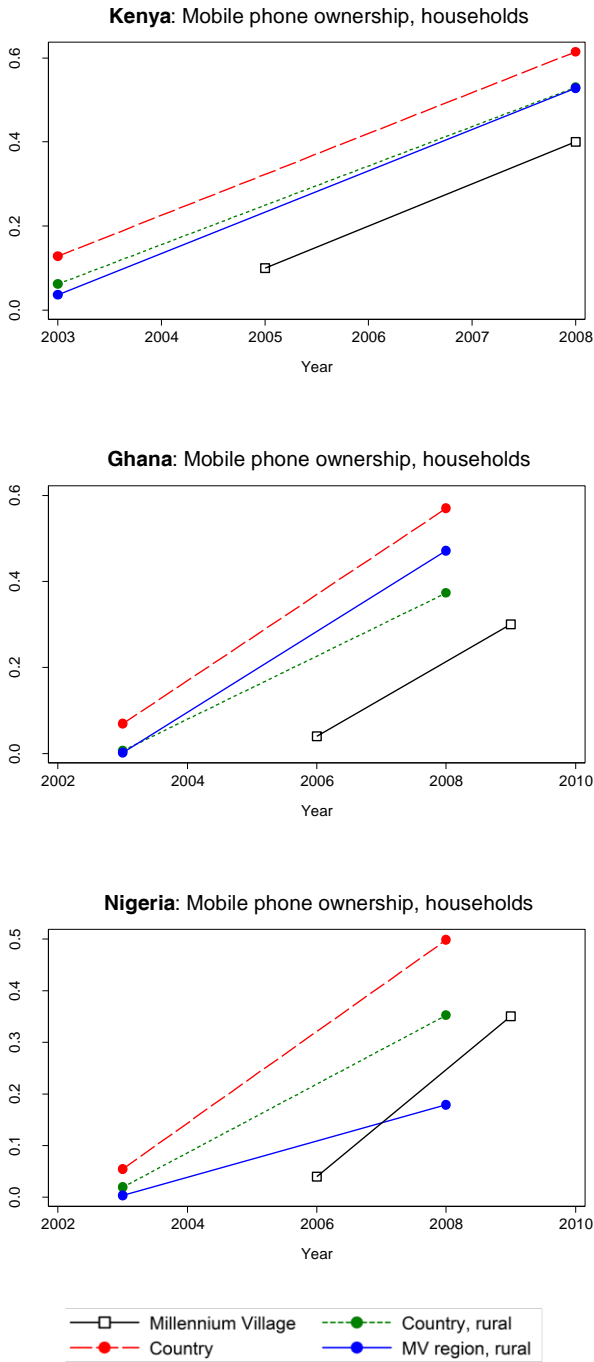
In June 2010, the project released its first public evaluation of the effects of the project (MVP 2010). The report compares values of several indicators at the time the project began to those three years into the program at five sites, describing the changes as “impacts” and “effects” of the project. A few months later, we produced a paper, later published in the *Journal of Development Effectiveness*, which critiqued the initial report and overall evaluation strategy (Clemens and Demombynes 2011). Our paper compared the changes at the sites to broader trends in the countries where the MV are found. For the project sites in Kenya, Ghana, and Nigeria, Figure 1 compares trends in mobile phone ownership with trends nationally, in rural areas of those countries, and in rural areas of the regions where the MV sites are located.

As the MV report notes, mobile phone ownership rose at the project sites. But it increased just as fast in areas untouched by the project, as the mobile revolution swept across the continent. This suggests that it would have risen at the MV sites with or without the project. Consequently, it is incorrect to view the rise as an “impact” of the project. Similar problems apply to all of the claims of impacts made in the 2010 MV report.

Returning to the notation used in the previous section, the before-and-after comparison used in the Millennium Villages report presented $Y_i(1)$ —the changes observed in each village—as the impacts of the project. But this would only be the impact of the project under the assumption that $Y_i(0)$ equals zero. Under this assumption, if the MVP had not been introduced, nothing would have changed: mobile phone ownership, child nutrition, access to clean water, etc. would have remained at the same levels as they were pre-MVP. The before-and-after approach to impact evaluation is widely recognized to be invalid because it requires this extreme and generally incorrect assumption that $Y_i(0)$ is zero.

² See Clemens and Demombynes (2011a) for a longer discussion of the experience and criticisms of IRD projects.

Figure 1: Fraction of Households that Own a Mobile Phone



2003 numbers include both mobile and landline phones, and are thus an upper bound on mobile phone ownership. All other years are mobiles only.
 Sources: MVP (2010) for Millennium Villages sites and authors' analysis of Demographic and Health Survey data for all other points.

Notes: the 2003 numbers include both mobile and landline phones, and are thus an upper bound on mobile phone ownership. All other years are mobiles only.

Our approach was to compare trends at the MV sites to trends regionally and nationally, in a version of what is called “differences-in-differences.” With this method, the trend for another location is used as a stand-in and assumed to match the counterfactual trend $Y_i(0)$. While this is unlikely to be precisely true, it is almost certainly a better approximation than the zero-change assumption.

Those comparisons show that for most indicators, conditions were improving regionally and nationally in the countries where the MV are located, as they have across much of sub-Saharan Africa in the last decade, for reasons that have no connection with the MVP.

As a result, our difference-in difference analysis suggested that the true impact of the MVP intervention, averaged across all reported indicators and all sites, was roughly half of what the project had claimed based on the before-and-after comparison.

c. The controversy

The working paper launched a vigorous public debate.³ We wrote blog posts which described a visit to the Sauri, Kenya site (Demombynes 2010) and summarized the paper (Clemens 2010). The initial posts generated commentary by several economics bloggers as well as journalists.⁴ The Millennium Villages Project responded on their own blog two days later (Pronyk, McArthur, Singh, and Sachs 2010). Their principal argument was that the comparison to regional trends was misleading because there are many interventions going on outside the MVP site that resemble components of the MVP. Nine days later we countered with another blog post, explaining,

“Measuring the impact of the Project means asking this question: What happened at sites that received the Project’s package intervention, relative to what would have happened at those sites in the absence of the Project? ‘In the absence of the Project’ does not mean in the absence of any interventions whatsoever—it means what would have happened without that specific project.”
(Clemens and Demombynes 2010).

Evaluation expert David McKenzie (2011) agreed, and found the MVP’s stance “baffling.” He blogged,

³ The part of these interactions that occurred between October 2010 and February 2012 is summarized in Clemens (2012).

⁴ These included Chris Blattman (2010) as well as Tyler Cowen (2010) of Marginal Revolution on October 12, and Laura Freschi (2010) of AidWatch on October 15. A column regarding the paper was published in the Financial Times on October 15 (Harford 2010), and an article appeared in the Daily Nation, a leading Kenya paper (Kelley 2010). Links to much of the press and blog discussion discussed in this paper can be found at <https://sites.google.com/site/gdemombynes/millennium-villages-evaluation>

“The purpose of having these similar control communities is precisely to control for all the other stuff going on in these countries which could be causing changes in the Millennium Villages regardless of the impacts of the MVP.”

We discussed this controversy with researchers, practitioners, policymakers, and journalists who followed the debate through the blogosphere. The next March, the MV evaluation was the subject of a debate at Oxford University broadcast on the web (CSAE 2011) and the discussion mushroomed across blogs, media outlets, and academic journals including the *American Journal of Clinical Nutrition* (AJCN).⁵ Our paper was peer-reviewed and published in the *Journal of Development Effectiveness* (Clemens and Demombynes 2011a).

On May 8, 2012, the *Lancet* published an analysis of the effects of the project on child mortality, described by the MVP as the primary study outcome for the project (Pronyk et al. 2012). The paper found that

“the average rate of reduction of mortality in children younger than 5 years of age was three-times faster in Millennium Village sites than in the most recent 10-year national rural trends (7.8% vs. 2.6%).”

The paper quickly generated buzz in the development blogosphere,⁶ and in scientific journals: the leading journal *Nature* published, the next day, an editorial critiquing the MVP evaluation declaring, “With transparency comes trust” (*Nature* editorial 2012). In a co-authored blog post published a few days later (Demombynes and Prydz 2012) and a subsequent comment in the *Lancet* (Bump et al. 2012), we pointed out that both sides of the calculation were flawed. Child mortality was in fact falling *more slowly* at the Project sites than it was nationwide. The MVP acknowledged the errors in a blog post (Pronyk 2012), and the *Lancet* subsequently published a retraction. Jeffrey Sachs then reversed two years of denials that differences-in-differences was a valid evaluation method for the Millennium Villages. He said,

“Demombynes and Clemens were correct in both cases (AJCN and *Lancet*) that the outcomes in the Millennium Villages should be compared with national trends during comparable time periods. Using a longer time period for the national data understates the recent national advances” (Oransky 2012).

⁵ These include the World Bank’s Africa Can blog (Demombynes 2011a, 2011b; Clemens and Demombynes 2011b), a back-and-forth on the Guardian’s Poverty Matters blog (Clemens and Demombynes 2011c, Bunting 2011, and Sachs 2011), and a write-up in the *Economist* (2011) magazine, as well as a number of blog discussions including Haddad (2011). That year, MV researchers also published a study on child nutrition (Remans et al. 2011) at the MV sites. We wrote a critique (Clemens and Demombynes 2012), and they acknowledged our key point in a response (Remans et al. 2012).

⁶ These included reactions from prominent development bloggers Annie Feighery (2012), Matt Colin (2012), Lee Crawford (2012), and Tom Murphy (2012).

The retraction generated further blog discussion as well as coverage in the *Economist* (2012), *Nature* magazine (Butler 2012), and *Foreign Policy* (Starobin 2013).

The retraction happened within days of the *Lancet* paper's release. Many researchers, evaluation professionals, and specialist journalists sought comment from us as soon as the *Lancet* paper began to circulate. Indeed, editors at the *Lancet* and at *Nature* learned of some of the paper's flaws from blog posts. The online debate ensured that by the time the *Lancet* article appeared, there was already a well-developed and transparent conversation about the MVP evaluation and observers who understood the core issues. Unlike debates that occur by email at conference seminars, the full text of the debate is online for anyone to read. This discussion would certainly have been slower and less public in the absence of the blogosphere.

3. Lessons on the new transparency in impact evaluation

The MVP evaluation controversy offers a useful arena in which to weigh concerns raised by critics of the new wave of IE, suggesting a series of lessons on its value and limitations. All of the concerns we cite are legitimate in some settings; the MVP case illustrates *which* settings.

a. RCTs are only one of many forms of IE

Many observers have lamented the new focus on the use of Randomized Controlled Trials (RCTs). In an RCT, potential program beneficiaries are randomly allocated to a “treatment” group that receives the program and a “control” group that does not. This procedure, long common for medical research, has gained substantial traction in development. Critics correctly point out that effects of an intervention on subjects in a contrived experiment may differ from effects on the broader population from which those subjects were drawn (e.g. Heckman and Smith 1995, Deaton 2010).

These important concerns about interpreting the results of contrived experiments apply only to parts of the rigorous IE toolbox. Advocates of IE are often incorrectly characterized as believing that RCTs are the only valid form of impact evaluation and should be used for every project. This has been our experience in the MVP debate, as the program's staff has insisted that “[e]conomists like Clemens and Demombynes should stop believing that the alleviation of suffering needs to wait for their controlled cluster randomized trials” (Pronyk, McArthur, Singh, and Sachs 2010) and “[t]he critics complain that we are not following their preferred approach of randomized trials. Many of these critics are consultants in randomized trials and apparently believe that such a method should be used every time” (Niang and Begashaw 2013).

We have never suggested anything resembling these positions and doubt that even the fiercest advocates of RCTs think they are the only valid form of impact evaluation. Our own work analyzed the MVP's impact exclusively with a nonrandomized method, differences-in-

differences. We did discuss how other IE methods, including randomization, could be applied to the MVP, while stressing in the clearest terms that RCTs are but one way to make the counterfactual more transparent.⁷ Any project evaluation has to find the right rigor in estimating the counterfactual, which balances benefits against costs.

b. IE can answer important questions even for complex, multifaceted interventions

Many critics of IE observe that some complex policy interventions cannot be reduced to a control and treatment. Sometimes the interactions between numerous facets of the intervention determine the overall impact, and sometimes the treatment must evolve as implementers learn. IE tools are best equipped for assessing *what* a project's impact was, not *how* that impact was achieved (Easterly 2009, Deaton 2010). This might shift research toward smaller questions and shift policy toward simpler interventions (Ravallion 2009, 2012; Rodrik 2009; Rosenzweig 2012).

These concerns are valid and it is certainly easier for many IE methods to assess *what* impact was achieved than to assess *how*. The “what” questions, however, are very important. Pritchett, Samji, and Hammer (2013) distinguish two types of impact evaluation questions: funders' questions and implementers' questions. Funders must choose resource allocations across projects, sectors, and countries. They include ministers, development banks, and philanthropists. Allocating their resources efficiently requires some sort of measure of *what* each alternative can achieve. Implementers, in contrast, must learn iteratively which actions achieve the best results, in complex and often unpredictable ways. They must learn *how*.

The MVP episode illustrates how IE methods in today's literature are better adapted to funders' questions than to implementers' questions. Questions of critical interest to funders of the MVP include “Does this project achieve its stated goal of generating self-sustaining economic growth?” and “Does this project reduce child mortality more than other interventions we could have funded?” We showed that even simple methods of IE help provide answers to those questions for a highly complex and adaptive project.

Implementers' questions are less easily answered by IE, including our analysis. Implementers' questions include questions like “What mix of malaria control and fertilizer

⁷ In Clemens and Demombynes (2011a), we wrote [emphasis added], 1) “*Rigorous evaluation methods, including fully objective matching, policy discontinuities, and randomized treatment, among others*—though often criticized as methods only of interest to academic purists—offer precisely this objectivity and clarity to practitioners.” 2) “A rigorous impact evaluation of the MVPs can be conducted in the future. *Several methods of rigorous impact evaluation are available, including careful differences-in-differences, regression discontinuity designs, randomized assignment of treatment, and others...*” 3) “Effects of the intervention at the next 20 or so sites could easily be rigorously evaluated—by the method of randomized treatment among matched pairs that we suggest above *or by another equally rigorous method*—at relatively low cost.”

credit would be the most useful intervention for the people of Ruhira, Uganda?” Pritchett et al. discuss how iterative applications of IE methods, embedded within organizations in real time, can help implementers “crawl the design space” to form and answer some questions that arise in the day-to-day process of making a project successful.

c. IE is more than just an academic issue; it can be more ethical than foregoing IE

Some observers express ethical reservations about the application of IE methods. For example, Ravallion (2012) notes the ethical complexities of withholding treatments that are known to work, in the name of maintaining the validity of the control group. Jeffrey Sachs stated that the MVP would not collect detailed data in comparison communities because it was “ethically not possible to do an intensive intervention of measurement without interventions of actual process” (Sachs 2006).

There are likewise ethical complexities in foregoing rigorous and transparent impact evaluation. In its 2012 *Lancet* paper, the Project suggested that it had caused large declines in child mortality at its intervention sites. We revealed that those declines were tracking similar broad trends which Demombynes and Trommlerova (2012) attribute in large measure to increased coverage of beds of insecticide-treated nets (ITNs).

While the MVP intervention costs thousands of dollars per household,⁸ ITN production and distribution cost roughly \$10 per household (Givewell 2012). This raises the possibility that the same money spent on the MVP intervention, if directed to different projects focusing exclusively on malaria control, could have prevented large numbers of children from dying. No assessment of this possibility can be made without transparent and rigorous evaluation of the MVP. Ethical use of scarce aid resource, then, requires transparent and rigorous IE. This continues to be the case even when some aspects of the overall treatment—in this example, the bednets that were part of the MVP package—already rest on a base of impact evidence.

This highlights another shortcoming of how IE methods are often applied today: many impact evaluations use a counterfactual scenario of “no intervention.” But the relevant economic question is what the value of those resources is relative to their next-best use. A more relevant impact evaluation would ask not just “Does intervention X increase teacher

⁸ The MVP does not publish full costs for each project site, inclusive of all administrative and in-kind contributions from its various partners. For a single site among the 15 intervention sites, the West Mamprusi and Builsa districts Millennium Villages in northern Ghana, full costs have been published according to transparency regulations of that site’s principal funder—the UK Dept. for International Development. The full cost of the intervention at this site is US\$27.1 million, to affect “up to 30,000 people” (about 6,000 households) and “achieve substantial poverty reduction for up to 2,250 households” (DFID 2011). This is about US\$4,500 per household affected, and US\$12,000 per household experiencing “substantial poverty reduction”.

attendance relative to the state-of-nature?” but instead “Which of four competing uses of education budgets raise teacher attendance the most?”

d. IE cannot answer every question, but the questions it can answer are important

Many observers correctly note that IE methods are less well-suited to answer a range of important questions. First, most IE methods as now applied are better-suited to measuring average treatment effects across some group, rather than capturing how treatment effects differ among subgroups (Heckman and Smith 1995, Deaton 2010, Ravallion 2012, Pritchett et al. 2013). While exogenous subgroup analysis and cross-cutting designs are increasingly common, the complex logistics and costly sample-sizes required limit the ability of many evaluations to capture heterogeneous effects. This is an important limitation, but combining different IE methods from different points on the spectrum of rigor can help. In Clemens and Demombynes (2011a) we calculate differences-in-differences estimates of the MVP impact at each of three very different sites, as well as discussing how to more rigorously evaluate the overall impact of the project across all sites.

Second, a common concern about IE methods is external validity, i.e. whether internally valid results from one evaluation would be informative about results in a scaled-up version of the project (Rodrik 2009, Acemoğlu 2010). This is less of a concern in the case of the MVP, which has operated in a wide range of settings. Another type of external validity is a more serious concern for all evaluation of the MVP to date, including our own work: a larger version of any NGO project is likely to be run and implemented by a government (Ravallion 2012). Bold et al. (2013) show that an otherwise identical schooling intervention in Kenya has strong effects on children’s test scores when implemented by an NGO and no discernible effect when implemented by the national government.

A third limitation of IE studies is that they typically measure only partial equilibrium effect of the intervention, while in the real world economic general equilibrium effects and political reactions to an intervention’s proximate impact may offset or reinforce that effect (Easterly 2009; Ravallion 2009, 2012; Acemoğlu 2010). This does not suggest that partial equilibrium estimates of project impact are uninformative, but rather that they alone are insufficient. Rigorous evidence of local impact of a project like the MVP would need to be coupled with careful thinking about the broader impact of the project, and scale-up would preferably involve the kind of ongoing, iterative, embedded learning advocated by Pritchett et al. (2013). IE is a very powerful tool, but must always be one of many tools in policy formation.

e. The new transparency lies not just in methods, but also in materials and medium

IE methods at their core are tools for transparency. All IE methods are different ways to make it the counterfactual as explicit and credible—that is, as transparent—as possible. And that same transparency goal inextricably requires two other transparency tools: transparent

materials on which the methods work (particularly data, but also documentation), and transparent *media* to communicate, critique, and correct results in the never-ending iterations of scientific discourse. The MVP evaluation episode illustrates these as well.

Our work on the MVP evaluation was only possible due to the open data policy of the Demographic and Health Surveys (DHS) project. DHS surveys have been conducted in 89 countries, many at multiple points in time. The microdata of DHS surveys is available for free download by any researcher who completes a simple registration. These data were the basis of the analyses in Clemens and Demombynes (2011a), Demombynes and Prydz (2012), and Bump et al. (2012). Without the DHS policy of open-access data we could not have known or documented that the trends at the MVP sites often largely reflected national and regional trends.

Policies of broad data access similar to that of the DHS have proliferated, and the field has been moving towards a standard of having microdata from published studies be available within a reasonable time period.⁹ The MVP, in contrast, does not make available any microdata that could be used to check its claims of impact. When we formally requested access to the microdata underlying published work, we were told that we would have to wait 7–8 years.¹⁰ Such a policy is not designed to facilitate critical inquiry.

We also highlight the important role of new communications media in publicizing errors in the evaluation and creating pressure for the retraction of incorrect statements. The online discussion was closely followed in the field and helped the MVP evaluation become a case study on development course syllabi at several universities.¹¹ The debate at Oxford was organized in response to the online discussion. The UK aid agency DfID required an independent differences-in-differences evaluation of the new MVP site in Ghana (Masset et al. 2013), in part in response to the online discussion. The editors of the *Lancet* learned about errors in the 2012 paper from a blog post and were already under pressure to demand a retraction of the incorrect findings before they received a formal letter.¹² All of these

⁹ Demombynes (2012) argues for a rule that data collected with public funds should be available within 2-3 years of collection.

¹⁰ We send our request on March 31, 2011. The MVP responded, “The project is currently immersed in a 10 year evaluation, with the final site scheduled to finish its surveys in July 2016. It’ll certainly take us some time after that for our internal team to clean, analyze and publish the results from the evaluation.” Given that the project’s initial phase of data collection ending in 2009 only entered scientific publications by 2011 and 2012, we correspondingly add 2–3 years to the year 2016 to arrive at the implied release date of 2018–2019, that is, 7–8 years after our request.

¹¹ These include Swarthmore, American University, Columbia University, New York University, the University of Texas, and Tufts University.

¹² The blog post appeared on May 10. On May 14, *Lancet* editor Richard Horton [wrote](#) on Twitter, “On MVP: Talks continue”. We submitted our letter (Bump et al. 2012) to the *Lancet* on May 16th at 14:58 GMT. It was accepted for publication about 90 minutes later.

interactions would almost certainly have happened much more slowly or not at all if our paper had been issued just five years earlier—before the prominence of development bloggers like Chris Blattman and Tom Murphy and before the possibility for instant decentralized dissemination of blog posts via Twitter.

4. Conclusions

Properly assessing a project's impact requires transparency. It cannot be done without a defensible and explicit counterfactual; the definition of impact is always relative to some counterfactual. It cannot be done without reasonable ways for researchers' data to be studied by other researchers. And it cannot be done without an interactive process of external critique. All three of these goals have been served by recent trends in development economics. We show how errors in the Millennium Villages impact evaluation were corrected via a confluence of improved methods, better materials, and new media..

We draw five lessons from this episode. Four of these relate to the role of impact evaluation methods in fostering transparency. First, the right rigor for any given impact evaluation may or may not be a randomized trial. Often it is possible to take a big step up in rigor at almost no cost, as we showed by applying differences-in-differences methods and public data to MVP impact claims. Second, IE is more amenable to funders' questions than project managers' questions. Our impact evaluation results for the MVP are informative to funders considering whether Integrated Rural Development gives as much bang for each buck as other approaches but less informative to the decisions of day-to-day project operation. Third, evaluation carries an ethical imperative: aid funds are scarce, and spending them without diligence to direct them toward their best use is not responsible. Only careful assessment of impacts can identify their best use. Fourth, many of the big questions in development are out of reach of IE methods. IE methods' role is to bring greater transparency to assessing the effects of interventions that are going to happen in delimited contexts. Other research methods are and should be applied to other types of questions.

The fifth lesson is that two other small revolutions have reshaped evaluation. The new transparency in impact evaluation comes not just from methods but also from materials (especially data) and medium (blogs). Our critiques were only credible because the open data policy of the Demographic and Health Surveys allowed us to freely assess MVP impact claims informed by an independent source. And the online discussion surrounding this issue vastly raised the magnitude and speed of its impact on the research literature and policy practice. The possibility for democratic critiques, sometimes originating far beyond the limited readership of academic journals, open-access data, and rapid correction of errors brings a new dynamism and transparency to the application of new evaluation methods.

5. References

- Acemoglu, Daron. 2010. “[Theory, General Equilibrium, and Political Economy in Development Economics](#)”. *Journal of Economic Perspectives* 24 (3): 17–32.
- Blattman, Christopher. 2010. “Impact evaluations, good and bad”. Blog. October 12. Available at <http://chrisblattman.com/2010/10/12/impact-evaluations-good-and-bad> [Accessed August 12, 2013].
- Binswanger-Mkhize, Hans. 2011. “On the wrong path: CGIAR strategy and results framework.” Blog post. Available from: <http://hansvins.blogspot.com/2011/03/on-wrong-path-cgiar-strategy-and.html> [Accessed August 12, 2013].
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2013. “[Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education](#)”. CGD Working Paper 321. Washington, DC: Center for Global Development.
- Bump, Jesse B., Michael A. Clemens, Gabriel Demombynes, and Lawrence Haddad. “[Concerns about the Millennium Villages project report](#)”. *The Lancet* 379 (9830): 1945.
- Bunting, Madeleine. 2011. “Millennium Villages Project: does the 'big bang' approach work?” *Guardian Poverty Matters Blog*. October 10. Available at <http://www.theguardian.com/global-development/poverty-matters/2011/oct/10/millennium-villages-project-impact-questions> [Accessed August 12, 2013].
- Butler, Declan. 2012. “[Poverty project opens to scrutiny](#)”. *Nature* 486, 165–166.
- Clemens, Michael A. 2010. “When Rigorous Impact Evaluation Is Not a Luxury: Scrutinizing the Millennium Villages”. October 11, 2010. Views from the Center. Available at: <http://www.cgdev.org/blog/when-rigorous-impact-evaluation-not-luxury-scrutinizing-millennium-villages> [Accessed August 12, 2013].
- Clemens, Michael A. 2012. “[The Collision of Development Goals and Impact Evaluation](#)”, working paper prepared for the 12th AFD/EUDN Conference, Paris, March 26, 2012, *Evaluation and its Discontents: Do We Learn from Experience in Development?* Proceedings of the 9th AFD-EUDN Conference, pages 169–197. Published as “[L'affrontement entre les objectifs de développement et l'évaluation d'impact](#)”, *Revue d'économie du développement*, 26 (4): 175–205.
- Clemens, Michael A. and Gabriel Demombynes. 2010. “Evaluating the Millennium Villages: Reply to the MVP + Upcoming Seminar with Comments from Jeff Sachs.” *Africa Can ... End Poverty* blog. October 22. Available at <http://blogs.worldbank.org/africacan/evaluating-the-millennium-villages-reply-to-the-mvp-upcoming-seminar-with-comments-from-jeff-sachs> [Accessed August 12, 2013].
- Clemens, Michael A. and Gabriel Demombynes. 2011a. “[When does rigorous impact evaluation make a difference? The case of the Millennium Villages](#)”. *Journal of Development Effectiveness* 3 (3): 305–339.
- Clemens, Michael A. and Gabriel Demombynes. 2011b. “Millennium Villages Project continues to systematically overstate its effects”, *Africa Can... End Poverty* blog. February

4. Available at <http://blogs.worldbank.org/africacan/the-millennium-villages-project-continues-to-systematically-overstate-its-effects> [Accessed August 12, 2013].
- Clemens, Michael A. and Gabriel Demombynes. 2011c. “Millennium Villages Project needs proper evaluation”, *Guardian Poverty Matters* Blog. October 19. Available at <http://www.theguardian.com/global-development/poverty-matters/2011/oct/19/millennium-villages-project-proper-evaluation> [Accessed August 12, 2013].
- Clemens, Michael A. and Gabriel Demombynes. 2012. “[Multisector intervention to accelerate reductions in child stunting: an independent critique of scientific method](#)”. *American Journal of Clinical Nutrition* 95 (3): 774–775.
- Colin, Matt. 2012. “Sachs the rainmaker”. Aid Thoughts blog. Available at <http://aidthoughts.org/?p=3330> [Accessed August 12, 2013].
- Cowen, Tyler. 2010. “Evaluating the Millennium Villages”. *Marginal Revolution* blog. October 12. Available at <http://marginalrevolution.com/marginalrevolution/2010/10/evaluating-the-millennium-villages.html> [Accessed August 12, 2013].
- Crawford, Lee. 2012. “Millennium Villages: impact evaluation is almost besides the point”. *Roving Bandit* blog. May 14. Available at <http://www.rovingbandit.com/2012/05/millennium-villages-impact-evaluation.html> [Accessed August 12, 2013].
- CSAE. 2011. “Video: Live Web Broadcast and Recordings”. Center for the Study of African Economies 25th Anniversary Conference, Economic Development in Africa, March 20–22. Available at <http://www.csae.ox.ac.uk/conferences/2011-EdiA/video.html> [Accessed August 12, 2013].
- Deaton, Angus. 2010. “[Instruments, Randomization, and Learning about Development](#)”. *Journal of Economic Literature* 48: 424–455.
- Demombynes. 2010. “Evaluating the Millennium Villages.” *Africa Can ... End Poverty* blog. October 11. Available at <http://blogs.worldbank.org/africacan/evaluating-the-millennium-villages> [Accessed August 12, 2013].
- Demombynes, Gabriel. 2011a. “MVP evaluation session at Oxford”. *Africa Can... End Poverty* blog. March 21. Available at <http://blogs.worldbank.org/africacan/mvp-evaluation-session-at-oxford> [Accessed August 12, 2013].
- Demombynes, Gabriel 2011b. “The Oxford Millennium Villages Debate”, *Africa Can... End Poverty* blog. June 2. Available at <http://blogs.worldbank.org/africacan/the-oxford-millennium-villages-debate> [Accessed August 12, 2013].
- Demombynes, Gabriel. 2012. “Opening Up Microdata Access in Africa”. *Development Impact* blog. October 1. Available at <http://blogs.worldbank.org/impactevaluations/opening-up-microdata-access-in-africa> [Accessed August 12, 2013].
- Demombynes, Gabriel and Espen Beer Prydz. 2012. “The Millennium Villages Project Impacts on Child Mortality”. *Development Impact* blog. May 10. Available at <http://blogs.worldbank.org/impactevaluations/the-millennium-villages-project-impacts-on-child-mortality> [Accessed August 12, 2013].

- Demombynes, Gabriel and Sofia Karina Trommlerova. 2012. “[What Has Driven the Decline of Infant Mortality in Kenya?](#)” Policy Research Working Paper 6057. Washington, DC: World Bank.
- DfID. 2011. “Business Case: Millennium Village in Northern Ghana”. Available at <http://www.cgdev.org/doc/2012/BusinessCaseandSummary202483%282%29.pdf> [Accessed August 12, 2013].
- Easterly, William. 2009. “[Can the West Save Africa?](#)” *Journal of Economic Literature* 47 (2): 373–447.
- Economist, The. 2011. “Economics Focus: The Big Push Back”. December 3. Available at <http://www.economist.com/node/21541001> [Accessed August 12, 2013].
- Economist, The. 2012. “Millennium Bugs: Jeffrey Sachs and the Millennium Villages”. *Economist* magazine *Feast and Famine* blog. May 14. Available at <http://www.economist.com/blogs/feastandfamine/2012/05/jeffrey-sachs-and-millennium-villages> [Accessed August 12, 2013].
- Feighery, Annie. 2012. “What Have We Learned from the Millennium Villages Project?” *UN Dispatch* blog. May 9. Available at <http://www.undispatch.com/what-have-we-learned-from-the-millennium-villages-project> [August 12, 2013].
- Freschi, Laura. 2010. “Millennium Villages: don’t work, don’t know or don’t care?”. *Aidwatch* Blog. October 15. Available at <http://aidwatchers.com/2010/10/millennium-villages-dont-work-dont-know-or-dont-care> [Accessed August 12, 2013].
- Givewell. 2012. “Mass distribution of long-lasting insecticide-treated nets (LLINs)”. December. Available at <http://www.givewell.org/international/technical/programs/insecticide-treated-nets> [Accessed August 12, 2013].
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M.J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: World Bank.
- Haddad, Lawrence. 2011. “Jeff Sachs: LVP of the MVP?” *Development Horizons* blog. October 14. Available at <http://www.developmenthorizons.com/2011/10/jeff-sachs-lvp-of-mvp.html> [Accessed August 12, 2013].
- Harford, Tim. 2010. “Attested development.” *Financial Times*. October 15. Available at <http://www.ft.com/intl/cms/s/2/692bb448-d66b-11df-81f0-00144feabdc0.html#axzz2bkblwMmD> [Accessed August 12, 2013].
- Heckman, James J. and Jeffrey A. Smith. 1995. “[Assessing the case for social experiments](#)”. *Journal of Economic Perspectives* 9 (2): 85–110.
- Kelley, Kevin. 2010. “Analysts discredit UN’s millennium villages project.” *Daily Nation*. October 16. Available at <http://www.nation.co.ke/News/politics/Analysts%20discredit%20UN%20millennium%20villages%20project%20/-/1064/1034132/-/8c8y8b/-/> [Accessed August 12, 2013].
- Masset, Edoardo, Arnab Acharya, Chris Barnett, and Tony Dogbe. 2013. “[An impact evaluation design for the Millennium Villages Project in Northern Ghana](#)”. *Journal of Development Effectiveness*. 5 (2): 137–157.
- McKenzie, David. 2011. “Jeff Sachs, the Millennium Villages Project, and Misconceptions about Impact Evaluation”, *Development Impact* blog, October 19. Available at

- <http://blogs.worldbank.org/impac evaluations/jeff-sachs-the-millennium-villages-project-and-misconceptions-about-impact-evaluation> [Accessed August 12, 2013].
- Murphy, Tom. 2012. “Just how effective is the Millennium Village Project at reducing child deaths?” *Christian Science Monitor Africa Monitor* blog. May 23. Available at <http://www.csmonitor.com/World/Africa/Africa-Monitor/2012/0523/Just-how-effective-is-the-Millennium-Village-Project-at-reducing-child-deaths> [Accessed August 12, 2013].
- MVP. 2007. “Millennium Villages Project: Overview”. New York: The Earth Institute at Columbia University, Millennium Promise, and the United Nations Development Program. Available at: http://commentariolus.com/MVInfokit_rev16.pdf.
- MVP. 2010. *Harvests of Development: The Millennium Villages After Three Years*. New York: The Earth Institute at Columbia University and Millennium Promise. Available at http://ciesin.columbia.edu/binaries/web/global/news/2010/mvp_midterm_report.pdf
- Nature* editorial. 2012. “[With Transparency Comes Trust](#)”. *Nature* 485 (147), May 10.
- Niang, Amadou and Belay Begashaw. 2013. “Africa’s Leaders Embrace the Millennium Village Project—and Here’s Why”. *Millennium Villages* blog. July 1. Available at <http://www.millenniumvillages.org/field-notes/africa-s-leaders-embrace-the-millennium-village-project-and-here-s-why> [Accessed August 12, 2013].
- Oransky, Ivan. 2012. “Millennium Villages Project forced to correct Lancet paper on foreign aid as leader leaves team”. *Retraction Watch* blog, May 31. Available at <http://retractionwatch.wordpress.com/2012/05/31/millennium-villages-project-forced-to-correct-lancet-paper-on-foreign-aid-as-leader-leaves-team> [Accessed August 12, 2013].
- Pritchett, Lant, Salimah Samji, and Jeffrey Hammer. 2013. “[It’s All About MeE: Using Structured Experiential Learning \(“e”\) to Crawl the Design Space](#)”. CGD Working Paper 322. Washington, DC: Center for Global Development.
- Pronyk, Paul. 2012. “Millennium Villages Project Corrects Lancet Paper”, *Millennium Villages* blog. May 18. Available at <http://www.millenniumvillages.org/field-notes/millennium-villages-project-corrects-lancet-paper> [Accessed August 12, 2013].
- Pronyk, Paul, John McArthur, Prabhjot Singh and Jeffrey Sachs. 2010. “Evaluating the Millennium Villages: A response to Clemens and Demombynes.” October 13. *Millennium Villages* blog. Available at <http://www.millenniumvillages.org/field-notes/evaluating-the-millennium-villages-a-response-to-clemens-and-demombynes-2> [Accessed August 12, 2013].
- Ravallion, Martin. 2009. “[Should the Randomistas Rule?](#)” *Economist’s Voice* 6(2), February.
- Ravallion, Martin. 2012. “[Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo’s Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty](#)”. *Journal of Economic Literature* 50 (1): 103–114.
- Remans, Roseline, Paul M Pronyk, Jessica C Fanzo, Jiehua Chen, Cheryl A Palm, Bennett Nemser, Maria Muniz, Alex Radunsky, Alem Hadera Abay, Mouctar Coulibaly, Joseph Mensah-Homiah, Margaret Wagah, Xiaoyi An, Christine Mwaura, Eva Quintana, Marie-Andrée Somers, Pedro A Sánchez, Sonia E Sachs, John W McArthur, and Jeffrey D Sachs for the Millennium Villages Study Group. 2011. “[Multisector intervention to](#)

- [accelerate reductions in child stunting: an observational study from 9 sub-Saharan African countries](#)”, *American Journal of Clinical Nutrition*. 94 (6): 1632–1642.
- Remans, Roseline, Marie-Andrée Somers, Bennett Nemser, and Paul M Pronyk on behalf of the Millennium Villages Project Study Group. 2012. “[Reply to MA Clemens and G Demombynes](#)”. *American Journal of Clinical Nutrition*. 95 (3): 775–777.
- Rodrik, Dani. 2009. “[The New Development Economics: We Shall Experiment, But How Shall We Learn?](#)” in Jessica Cohen and William Easterly, eds., *What Works in Development? Thinking Big and Thinking Small*, Washington, DC: The Brookings Institution, pp. 24–47.
- Rosenzweig, Mark. 2012. “[Thinking Small: A Review of Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty by Abhijit Banerjee and Esther Duflo](#)”. *Journal of Economic Literature* 50 (1): 115–127.
- Sachs, Jeffrey. 2006. “The Millennium Villages Project: A New Approach to Ending Poverty”, speech at the Center for Global Development, Washington, DC, March 14. Transcript available at http://www.cgdev.org/doc/event%20docs/3.14.06_Sachs/Sachs%20transcript.pdf [Accessed August 12, 2013].
- Sachs, Jeffrey. 2011. “The Millennium Villages Project is working well”. *Guardian Poverty Matters Blog*. Available at <http://www.theguardian.com/global-development/poverty-matters/2011/oct/13/millennium-villages-project-working-well> [Accessed August 12, 2013].
- Sachs, Jeffrey, John McArthur, Guido Schmidt-Traub, Margaret Kruk, Chandrika Bahadur, Michael Faye, and Gordon McCord. 2004. “[Ending Africa’s Poverty Trap](#)”. *Brookings Papers on Economic Activity*. 2004 (1): 117–240.
- Sanchez, Pedro et al. 2007. “[The African Millennium Villages](#)”. *Proceedings of the National Academy of Sciences* 104 (43): 16775–16780.
- Starobin, Paul. 2013. “[Does it take a village?](#)” *Foreign Policy*, July/August.