

***Preliminary and Not for Quotation***

***Comments Welcome<sup>1</sup>***

*This Draft: March 21, 2011*

## **Designing, Implementing and Evaluating Complex, Goal-Oriented, Adaptive Interventions in the Millennium Villages**

*John W McArthur, Paul M Pronyk, Jeffrey D Sachs*

### **I. Introduction**

At the UN Millennium Summit in September 2000, world leaders adopted the Millennium Declaration, committing nations to a new global partnership to reduce extreme poverty and address pressing challenges of hunger, gender inequality, illiteracy, and disease [1]. The Declaration established a series of measurable targets for 2015, anchored in the ambition that extreme poverty in its many forms could be cut by at least half within a generation. During 2002-5, the UN Millennium Project drew from a global community of hundreds of scientists and practitioners to distill the evidence on how the goals could be achieved [2]. The project identified the evidence from a broad range of sources and disciplines and concluded that the scaled implementation of public investments in known technologies, in the context of open, well governed, and market-based economies, could enable the achievement of the Millennium Development Goals (MDGs). Thus the MDGs were framed as a matter of implementation of a well-balanced mixed economy, involving private-led economic growth and public investments in critical areas, including health, education, infrastructure, agriculture, and environmental sustainability. As summarized in several global agreements, including the Monterrey Consensus, the implementation would be country led, market based, and supported by the international community through trade, technology, and development assistance.

Working alongside other initiatives, the UN Millennium Project made a series of policy recommendations to the global community, the majority of which were adopted at the 2005 G8 Summit in Gleneagles and at the 2005 UN World Summit in New York. The recommendations emphasized a several practical priorities, ranging from “quick wins” in malaria control to long-term investments in agriculture and infrastructure. At these summits, world leaders made major commitments to increase official development assistance (ODA) to implement basic programs, particularly in sub-Saharan Africa, with a pledge to make available ODA of approximately \$85-100 per African by 2010 en route to much

---

<sup>1</sup> Please send comments to [ppronyk@ei.columbia.edu](mailto:ppronyk@ei.columbia.edu).

further increases by 2015.<sup>2</sup> These financial commitments underpinned a major policy breakthrough which stated that every developing country would be supported to implement a national strategy ambitious enough to achieve the MDGs.

Amidst the progress embodied in these international agreements, the world at the end of 2005 still confronted a basic implementation gap. There was as yet no detailed year-by-year timetable for the 2005 financial commitments to be brought to fruition, and there were many practical questions about the training, management, monitoring, information systems, and financing needed to scale up the relevant public investments in the context of the mixed public-private strategy. Many specific things needed to be brought from basic concepts to implementation success, including the design of effective delivery systems that could reach all the way from the global community to remote local villages. This raised particular challenges for the impoverished villages of Sub-Saharan Africa, home to hundreds of millions of the world's poorest people.

The Millennium Villages Project (MVP) was therefore launched during 2005-6 to help design, measure, and scale up effective delivery systems to help rural populations achieve the MDGs on an integrated basis. The aim was to address a major implementation challenge: how could an integrated set of public investments and the international policy and financing agreements be translated rapidly and at scale into effective local investments in priority areas. Thus the core task motivating the Millennium Villages was framed as follows:

To create a community-based implementation system to achieve the MDGs in impoverished rural African populations across many distinct agro-ecological zones while operating within the per capita budget envelope as defined by the international ODA targets and national budget capabilities.

The MVP focuses on regions that are far off course from achieving the MDGs. Unfortunately, that still includes much of rural Africa. The goal was and is to identify an effective local implementation system that can operate in a coordinated manner across several interconnected sectors, notably: agriculture, health, education, local infrastructure, business development, and the environment. The budget model is based on a reasonable implementation of national budgets and global development assistance standards; and the technological choices are based on proven best practices in the key areas. The implementation model is designed and refined on an ongoing basis in the course of the MVP itself, through learning by doing in each location, drawing on extensive development knowledge, national and local institutions, and evolving tools such as mobile-phone based systems design.

It is important to clarify that the Project was not launched to assess whether the MDGs could or should be achieved through a "one village at a time" delivery model. Instead it was launched to examine how, and whether, a contextually appropriate set of proven interventions could be systematically

---

<sup>2</sup> The calculation per African is based on a commitment to reach at least \$50 billion in 2005 dollars by 2010. Dividing by a target population in need of approximately 500-600 million people works out to \$83-100 per person.

implemented in impoverished and remote rural settings in a manner consistent with achieving an objective set of MDG targets. In effect this is a test of whether a series of community-based (“last mile”) service delivery challenges can be addressed in an integrated manner and within a strict and transparent budget constraint. Despite broad global policy commitments to achieve the MDGs, the MVP is one of the only, and perhaps the only, initiative that is systematically introducing a community-based system for comprehensive MDG implementation across a highly diverse range of agro-ecological settings.

As a complex intervention program that works across multiple sectors and in numerous contexts, the Project poses a host of important challenges to evaluation. Recently, skeptics have criticized the Project and its evaluation methodology as lacking in rigor and objectivity [3], with recent comments on the World Bank’s website suggesting the Project has been “systematically overstating its effects” [4]. This Commentary seeks to respond to these issues, and clarify a number of fundamental points about the Project’s inception, design, and goals. It does so while clarifying certain key issues regarding the most appropriate methods for evaluating complex interventions to accelerate access to the MDG targets.

## **II. Overview of the Millennium Villages Project**

The Millennium Villages is a ten-year project running from 2005-6 to 2015, the deadline year for the Millennium Development Goals (MDGs). The *primary aim* of the project is to achieve the clear quantifiable endpoints in relation to the Millennium Development Goals in all Project sites, as a contribution to the fulfillment of the MDGs more generally. These site-based level targets for 2015 are a general translation of the MDG targets to the conditions of poor, rural subsistence farming communities in sub-Saharan Africa. They are outlined in Table 1.

The Project operates in a diversity of agro-ecological zones, representing a range of challenges to income, food production, disease ecology, infrastructure, and health system development. First initiated in Kenya and Ethiopia in 2005 and then launched at scale in 2006, the Project currently reaches approximately half a million people across 12 sites in 10 countries. The MVP is a partnership of the Earth Institute at Columbia University and Millennium Promise, an international NGO. The host communities and governments implement the project with the leadership and staffing of local teams in all cases. Many of the critical support operations of the MVP are carried out by two regional offices: the MDG Center for East and Southern Africa in Nairobi, Kenya, and the MDG Center for West Africa in Bamako, Mali. The United Nations Development Program (UNDP) has provided administrative support in Phase I (2006 to mid-2011). An array of other international organizations contribute commodities and technical inputs to the project, including UNAIDS, UNICEF, and the World Food Program, as do a variety of major international companies, including Agrium, Ericsson, JM Eagle, Mosaic, Novartis, and Sumitomo Chemical.

**Table 1: Millennium Villages Project MDG-based targets**

**MDG 1:**

Poverty metric (World Bank poverty line of \$1.25 per day in \$2005 PPP) < 25%  
Stunting among children under 5 years old < 20%  
Wasting among children under 5 years old < 5%

**MDG 2:**

Net primary attendance >90%

**MDG 3:**

Gender parity in primary education > 0.9

**MDG 4:**

Child mortality <40/1000  
Measles immunization & vitamin A supplementation > 90%

**MDG 5:**

Maternal mortality < 150/100 000  
Skilled birth attendance/Institutional delivery >70%  
Modern contraception use > 40%

**MDG 6:**

Infant HIV infections from mother to child < 5%  
ARV coverage >85%  
TB treatment success >85%  
Malaria prevalence < 5%

**MDG 7:**

Access to improved drinking water > 90%  
Access to Improved sanitation > 75%

**MDG 8:**

Universal access to essential medicines (ARVs for HIV/AIDS, and ACTs for malaria)  
Broadband connectivity in all clusters

To achieve this aim, the Project has been designed around several key **organizing principles**.

- The MVP emphasizes the deployment of low-cost interventions that have been proven in earlier studies, often through randomized control trials, in areas of smallholder agriculture, public health, primary education, local infrastructure, and business development. The purpose is not to re-test the individual interventions, but rather to demonstrate the feasibility of their joint implementation according to a pre-established budget and timeline.
- The MVP aims to design and document systems of interventions that can be delivered effectively with local resources plus the external aid envelope promised by the donor countries at Monterrey in 2002 and Gleneagles in 2005.
- The MVPs emphasize community-based management of the requisite systems, as well as the feasibility of adapting and tailoring these systems to meet diverse local needs across Sub-Saharan Africa's major agro-ecological zones.
- Local teams are the primary managers of successful implementation of the MVP systems.
- The MVP partnership of the EI and the MP provide scientific, technical, and management support.

The MVP is generating a range of **products** that provide a bridge to national and Africa-wide MDG scaling up in several ways.

- The MVP is designing, documenting, and disseminating a range of implementation tools that can be widely applied. These include community-based information and management systems, training manuals and courses, computer and mobile-phone-based applications for monitoring and enhancing intervention coverage in rural areas, standards for well-functioning institutional designs, and other "institutional and management capital".
- The MVP operates within a consistent budget envelope and is carefully measuring the costs of various delivery systems and interventions, for use in needs assessments and budgeting.
- The MVP is developing a thorough evidence base of the institutional capacities and needs of communities in a diverse array of agro-ecological conditions.
- The MVP is fostering an Africa-wide team of expertise that is working with local, national, regional, and AU officials on integrated approaches to MDG scale up, informed by lessons learned in the MVPs.

At the same time, it is worth clarifying what the MVP is **not** trying to do:

- The MVP is not testing a rigid protocol for implementing MDG-based systems. The emphasis is community and project learning, design, and local context. Through careful documentation and codification of these learning mechanisms, the MVP will support broader, ecologically attuned programs aimed at scaling up.
- The MVP is not claiming to provide the only model for achieving the MDGs. It is aiming to provide a workable model that fits the relevant budget constraints and that is *adequate* (sufficient) to achieve the MDGs by 2015. It does not claim uniqueness or even optimality. Yet given the pervasive shortfalls of the MDGs throughout rural Africa, adequacy of the MVs in meeting the MDGs is of great practical and conceptual importance. We are not aware of any other initiative that is rigorously documenting the processes, systems, and budgets needed to achieve the MDGs at the scale of hundreds of thousands of people across distinct ecological-zones.

### **III. Evaluating the Millennium Villages**

Evaluation is vital – both for strengthening the link between good science and sound policy, as well as to ensure public confidence in how limited resources are deployed. With hundreds of millions of people still living in extreme poverty, there is a pressing need to better understand how rapid gains can be achieved to maximize the impact of every dollar or pound spent.

#### *Study design options*

There are a number of methodological choices for evaluating complex interventions depending on the kind of inference being made, as well as the degree of confidence that observed changes are a result of a particular program [7].

An *adequacy* assessment compares the performance of the intervention package (such as the MVP) against a fixed set of goals, such as the MDGs. The purpose is to demonstrate that the package is adequate to meet the designated objectives. There is no control group, since the standard of assessment is the outcome of the project compared with the goals of the project.

A *plausibility* assessment takes causal inference a step further, to make a plausible case that the observed changes are largely the result of the intervention package rather than of external or confounding forces. Confounding effects might include factors such as general economic changes, climate, or external shocks – all of which might influence key outcomes. A plausibility assessment is generally carried out with two sets of methods. First, the project identifies the impact pathways that link the interventions with the outcomes, and then assesses the impact pathways through detailed

measurements, including the timing of interventions and outcomes [8,9,10]. Second, the project outcomes are compared against a set of reference data, which might include a designated comparison site, a neighboring district or sub-district, the country-wide average for rural areas, and/or some other meaningful comparison region.

A *probability* assessment uses a randomized trial approach to rule out the effects of confounding factors, bias, or chance. This design is particularly appropriate for assessing the efficacy of unproven interventions and for those that can be delivered in a consistent manner across treatment areas. It is suitable where a relatively large number of units can be randomized among treatment and comparison groups, and where the comparison groups are relatively untouched by the intervention being assessed [11].

### *The MVP evaluation platform*

The MV evaluation framework is summarized below, and draws from a range of assessment methods to respond to research questions outlined in Table 2.

**Table 2: Key research questions of the Millennium Villages**

**Adequacy**

- Can the MDGs be achieved within a fixed budget envelope, over what time period (adequacy assessment)?

**Plausibility**

- Is there a plausible relationship between intervention delivery, levels of coverage and observed effects?
- Are synergies achieved from working simultaneously across multiple sectors? Are effects greater than those estimated by single sector effectiveness trials?
- What are adequate intervention doses and levels of coverage required to observe population level effects?

**Process Design**

- How is the MV model adapted to meet the needs of local contexts?
- What is an adequate timing and sequence of interventions?
- What are the barriers and facilitators to implementation in each setting?

**Costs**

- What is the cost of implementation of the overall package relative to the \$120 threshold, and what are the returns on investment according to various metrics (health, social, or dollar returns, for example)?

**Scale-up Tools**

- Which systems (m-health, soil analysis, Community Health Worker training, etc.) can be standardized for mass scaling up?
- Which systems are susceptible to ICT solutions?
- What are the most effective means for dissemination of effective MDG systems (training, government consultations, handbooks, conferences, etc.)?

### *Adequacy and Plausibility*

The core of MVP M&E includes both adequacy and plausibility assessments. *The MVP is not a Probability Assessment.* There are a number of reasons why the project rejected the use of a randomized trial model:

- The intervention approach includes a bundle of dozens of interventions, applied with locally defined variation, rather than a single intervention or fixed set of interventions.
- Most of the interventions are well known, and previously demonstrated, so it is not a priority to identify whether they work individually
- A large number of the interventions are carried out at community scale (e.g. building a school, a clinic or a road), not at individual or household scale, so that the number of independent observations is necessarily small (i.e. scaled by the number of communities, not the population within them)
- The effective modalities of *concurrent delivery* of multiple known interventions were not known at the start of the project, as there is considerable learning-by-doing in defining integrated sets of intervention packages
- Each site within the project is unique because of its distinctive agro-ecological characteristics (which also affect health, infrastructure, livelihoods, hazards, infrastructure, and many other characteristics)
- Amidst the practicalities of joint decision-making with local governments and time-bound nature of the initiative, a uniform process of random selection across 10 countries was considered unfeasible at the project's outset.
- In practical terms the absolute value of financial resources required to do large scale-randomization of multi-sector intervention packages within an agro-ecological zone for clusters of 30,000 to 50,000 people at a time would require budget envelopes on the order of those only made available by major international financial institutions, and would provide unclear incremental information for the reasons just stated. To illustrate, a set of 10 pairs of comparison populations of 30,000 people within the same agro-ecological zone would on its own require more than \$20 million per year and \$100 million over five years.<sup>3</sup> Instead, the MVP undertook a single cluster per agro-ecological zone. (Recognizing the above analytical limitations, the MVP, which is primarily financed by private philanthropists, would be pleased to pursue such a collaborative research exercise if an institution like the World Bank wants to make the corresponding investment.)

---

<sup>3</sup> Multiply 10 treatment sites by 30,000 per site, for a total of 300,000 people. Multiply that by a \$60 annual per capita intervention cost, for a total of \$18 million per year. Add \$10 per capita for annual measurement and coordination with comparison sites, summing to \$3 million per year. Thus \$18 million plus \$3 million = \$21 million per year. A five year assessment costs \$105 million. A ten year assessment costs \$210 million (in constant dollars).



These types of challenges are increasingly common in complex intervention studies. To that end a recent *Lancet* review suggests “a reductionist approach to evaluation based on isolation of program effects is no longer appropriate for scaling up of initiatives to reach the MDGs in most low-income countries”[11]. Others suggest an emphasis on randomized designs is both overly narrow and unlikely, on its own, to provide useful insights regarding “what works” about development processes [12].

Since randomized trials are not the most appropriate approach for the MVP, the adequacy and plausibility standards are assessed by monitoring population-level progress towards a series of pre-determined MDG targets, as outlined in Table 1. The plausibility that observed changes are a result of the intervention package is being assessed as follows:

- By comparing observed changes to a specific comparison village site established by the Project
- By comparing observed changes to trends in levels of intervention coverage and MDG outcomes in other reference data, notably other national and sub-national information and data sets
- Through measuring the consistency of changes across diverse project sites, where population level confounding factors are likely to differ
- By mapping the observed changes to pre-determined impact pathways
- By detailed monitoring of the timing and sequence of interventions in relation to observed shifts in levels of coverage and changes in outcome variables

In other words, the Project takes a multi-faceted evaluation approach that relies not only on comparison/control groups but also on broader reference data and on dosage assessments, pathway analysis, time series assessments and implementation research.

#### *Site selection*

The original Millennium Village site selection was driven by a range of objective criteria, with the aim of maximizing the learning potential of the project, both by identifying very poor and difficult locations, and by identifying locations across diverse agro-ecological settings. Sites were chosen purposively to represent over 95% of the agro-ecological zones on the continent – reflecting a variety of systems-level challenges, disease profiles, and baseline levels of infrastructure and capacity. The number of sites was a product of the available resources to support a concerted multi-sector approach at the total cost threshold of \$120 per capita, including \$60 of external MVP support alongside \$60 from local sources.

Within each agro-ecological zone, contiguous village clusters of approximately 30,000 to 60,000 people were chosen from ‘hunger hot-spots’ as designated by the UN Millennium Hunger Task Force - where at least 20% of children under the age of five were malnourished [5] [6]. Final selection was “needs-based” and made together by government and the Project. In nearly all sites, by virtue of being “hunger hotspots,” the baseline levels of poverty and access to basic services were below national rural levels, making it highly unlikely sites were systematically advantaged from the outset.

*Comparison sites and data collection*

As part of the reference data for the plausibility assessment, a set of matched set of comparison villages (CVs) were chosen reference points against which to compare changes in MDG-related indicators across the Millennium Village (MV) sites. These were introduced in year 3 of the project. These villages do not provide a true control group, as governments, NGOs and other development partners are currently involved in scaling up many of the same interventions contained in the Millennium Villages package, though typically with a lower and varying dosage, i.e., much less consistency and intensity compared with the MVs. Nonetheless, the CVs provide some useful information on the types of MV-type activities being implemented in other rural areas in the country, while providing additional information on country-level confounding factors such as climate, economic change, political instability and so forth. Data is collected at the same points in time as MV project sites, allowing for more robust site-specific and local area comparisons.

In addition to the CVs, the project also collects extensive data from the nearby districts, as well as from national level data for rural areas. We do not consider the MVs and the CVs to constitute a probability assessment, though we do make statistical comparisons across the pairwise samples. This analysis is limited by the fact that the CVs are not true control groups, as mentioned, and the fact that each MV site is a distinct agro-ecological zone, with different response rates to different combinations of inputs. For example, each site has its own agronomic challenges and its own epidemiological profile. It is thus not entirely satisfactory to pool all of the MVs into a single statistical test of 10x2 sites. While we do so for some tests of the project, we acknowledge and underscore the inevitable limitations of such a procedure.

*Process evaluation*

Qualitative process evaluation or implementation research involves systematic interviews with site teams, government officials, local partners and community members. This examines the timing and sequence of interventions, local adaptations to the MV model, and key barriers and facilitators to implementation. This component helps to answer design-level questions such as: “How can community health workers armed with a cell-phone effectively diagnose and treat malaria at the household level?” or “What are the most appropriate mechanisms to provide small-scale loans for agriculture?”

*Costing assessment:*

A further economic costing module documents project spending (real and in-kind) by year, sector and stakeholder, to provide detailed estimates of the costs of various interventions and systems, information that is vital for scaling up and that is very difficult to obtain other than through implementation projects such as the MVP. This component will assess whether the MDGs can be achieved for the projected \$120 per capita total investment envelope. When paired with micro-economic data on household and village level shifts in economic and social well-being, these data will also assist in assessing returns on investment.

*Transparency and project oversight*

To enhance the transparency of the evaluation, and subject the methodology to scientific scrutiny uncommon to the evaluation of most development interventions, a study protocol has undergone external scientific peer-review and been registered with *The Lancet* (Protocol number 09PRT-8648). In addition, the trial has been registered with the United States National Institutes of Health at ClinicalTrials.gov (protocol number NCT01125618). Finally, all study procedures are subject to oversight and annual updates by Institutional Review Boards at Columbia University and 10 host-country governments.

**IV. Response to specific critiques**

The remainder of this Commentary will respond to some specific criticisms of the project and its' evaluation. Many of these focus on a June 2010 progress report entitled *Harvests of Development (HOD)* which was published in the lead up to the September UN MDG Summit. This has been repeatedly and inaccurately described as an MVP evaluation report, resulting in a number of incorrect assertions that might confuse readers on substantive points of agreement and disagreement.

HOD reported observed changes after three years of program exposure among a dozen MDG-related indicators, and was cautious to avoid presenting raw data or statistical testing so as to not jeopardize future publications. The first page of the report states the following:

This report highlights the early results after three years of implementation across five initial Millennium Village sites in Ghana, Kenya, Malawi, Nigeria, and Uganda. Progress toward achieving the MDGs are derived from recently completed mid-term (year three) surveys. All data contained in this report compare baseline values to year three assessments, among a sample of several hundred households across each cluster. (...)

Further scientific results, including comparisons with other villages, will be published later this year, including in peer-reviewed scientific literature. *We therefore emphasize the provisional nature of the results presented here, both in the sense that they are after only the third year of a ten-year project, and in that they represent only part of the third-year evaluation underway this year. We are presenting these partial results now in order to foster a better public understanding of the Project and its potential to help reduce extreme poverty, hunger, and disease in rural Africa. We hope that this report contributes to the public discussion in the lead-up to the MDG Summit in September 2010.* [p.4, italics added.]

In short, *HOD* was not a formal evaluation of the project, and treating as such is incorrect and ignores the clear point on the first page emphasizing the very preliminary nature of the findings.

- **Should the evaluation design have employed cluster randomized trial?**

A large and growing body of literature assumes that randomized control trials are the only method for developing a rigorous assessment of development programs. We have already explained the specific features of a complex intervention program like the MVP that make cluster randomized trials inappropriate. This results from the fact that the interventions are at community scale, are complex, are essentially one site per agro-ecological zone, and are without true control regions. Thus, we do not aim for a formal probability assessment, but only for an adequacy and plausibility assessment.

- **Are the Millennium Villages overstating their effects?**

Another argument is that the significant observed changes presented in the *HOD* report should be downplayed due to broader secular improvements generated by mapping national and sub-national trends from Demographic and Health Surveillance data. A few important clarifications are in order here. One is that even in criticizing the MVP's progress report, the critique itself asserts that the MVP is making roughly twice as much progress as other locations. The MVP is on track to achieve the MDGs in almost all or all sites, whereas most of rural Africa is falling short of most or all of the MDGs.

The MVP of course recognizes and celebrates that some progress on the MDGs is indeed occurring throughout much of rural Africa, largely because the same kinds of interventions (e.g. bed nets, medicines, mobile-based health care, etc.) are being taken up more widely. This uptake is the result, in part, of the international adoption of the same recommendations put forward by the UN Millennium Project and its partners in 2005, and the subsequent uptake of these interventions in dozens of low-income countries in Africa and other regions. In many cases it is also the result of policy spillovers from the MVP. Unfortunately, however, the interventions outside of the MVs are generally not reaching the scale and intensity that is needed, so that the results of the MVs outpace the results in rural Africa more generally.

- **Do MVP statements on cell phone usage suggest systematic overstatements?**

A paper and multiple blog posts by Clemens and Demombynes make much of a text box on p.91 of HoD, which includes the word "impact" alongside metrics on mobile phone usage. First, it should be clarified that the use of the word impact was in this instance used in an everyday sense for a general audience, and did not mean to imply a technical meaning of change relative to an underlying trajectory or control. As we have said to Clemens and Demombynes, we accept that a better word should have been chosen. MVP scientists are well aware of the fast growing progress of connectivity around Africa, and indeed speak frequently in public venues about that increase in connectivity.

Second, and perhaps paradoxically, there is an error in assuming that all connectivity growth in the MVP has been of the same character as in other regions, since the project has had a major partnership with Ericsson and local mobile providers to ensure that MVP communities received connectivity on an accelerated basis, i.e., ahead of the market. This in turn flags two important issues. One is that more careful case research of the project would show that several mobile towers were indeed installed on a pre-market timetable, including those to eliminate connectivity gaps in coverage within the cluster. The MVP can point to many specific health clinics and schools that thereby received cell phone connectivity as a result of the project well ahead of a market rollout, and that accelerated connectivity then enabled the introduction of specific interventions (such as mobile-health technologies and school connectivity).

- **Should one wait 15 years for long-term follow-up studies to be conducted prior to scaling up interventions?**

Some critics suggest that efforts to take interventions to scale should wait at least 15 years until evidence of long-term effects can be proven and sustained. Part of this recommendation is likely based on a misunderstanding of what the MVP's community level delivery systems are aiming to demonstrate and achieve. Part of it is likely based on inadvertently overlooking the fact that there exists a deep, proven record of what needs to be done. The key challenges remain the design of effective systems with parallel financing for implementation.

The global history of the HIV pandemic paints a compelling picture of the potential consequence of such delays. In 2000 roughly 36 million people had been infected by the HIV virus, which had already resulted in a total of 20 million deaths – with most of those in sub-Saharan Africa. Highly active antiretroviral therapy (HAART) had been successfully introduced to rich country markets in 1996, resulting in durable reductions in viral load and major survival gains. However, as late as 2001, there was no clear protocol for using them in low-income countries.

Waiting 15 years to examine whether the results of randomized trials could be sustained before treatment systems were scaled up, many more would have died. Fortunately, as recommended by the Commission on Macroeconomics and Health, the Harvard Consensus Statement on HIV/AIDS, and other groups, and based on the scientific evidence available to the Commission and to the Harvard signatories, a general scaling up began in 2001. A key step was the launch of the Global Fund to Fight AIDS, TB, and Malaria in 2001 and the US President's Emergency Program for AIDS Relief in 2003. These institutions were launched without a randomized control trial among different global institutional models for donor coordination and country-led service delivery. Instead, they were launched on a premise of achieving delivery targets and ongoing learning by doing and international peer review. Many important design and efficiency lessons have been learned in the course of the ensuing decade, informed by advances in diagnosis, treatment and prevention. The rapid scaling of HAART implementation in some African countries resulted in reversals in previous declines in life expectancy [13]. In those that delayed, countless lives were lost [14]. Similar analogies of rapid scaling

up can be drawn to bed nets, high-yield seed varieties, micro-finance, fertilizers, emergency obstetrical care, and many other important proven interventions.

## **V. Conclusions**

A foremost challenge of global development is to find ways of implementing proven interventions at a significant scale, with systematic monitoring, evaluation, budgets, management, scale-up and feedback. Systems are constantly improving thanks to careful ongoing research and debate, rather than misleadingly binary framings of whether or not to proceed. Those lessons merit broader application across the specialized fields of health, agriculture, infrastructure, education, and business development, which is why the Millennium Villages can be important for thinking through the balance of research and system design for integrated service delivery across a range of disparate locations on a real time basis at a time when more than half a billion people still live on less than \$1/day in extremely poor rural villages. Decades of research on countless areas of agriculture, health, education, infrastructure, and business development have shown many crucial and scalable lessons of what works. The key now is to apply this knowledge systematically. A systems approach is required for that purpose, and the MVP provides one important learning option for such an approach.

## References

1. United-Nations-General-Assembly, *United Nations Millennium Declaration*, in 55/2. 2000.
2. UN Millennium Project, *Investing in development: a practical plan to achieve the Millennium Development Goals. Overview*. 2005. London.
3. Clemens, M. and G. Demombynes, *When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages*. 2010, Center for Global Development, Working Paper 225: Washington.
4. Demombynes, G., *Millennium Villages Project continues to systematically overstate its effects*, in <http://blogs.worldbank.org/african/millennium-villages-project-continues-to-systematically-overstate-its-effects>, S. Devarajan, Editor. 2011, World Bank: Washington.
5. Sanchez, P., et al., *Halving hunger: can it be done*. 2005, UN Millennium Project Task force on Hunger: New York.
6. Dixon, J., A. Gulliver, and D. Gibbon, *Farming Systems and Poverty*. 2001, Rome: FAO.
7. Habicht, J., C. Victora, and J. Vaughan, *Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact*. *International Journal of Epidemiology*, 1999. **28**: p. 10-18.
8. Bryce, J., et al., *Programmatic pathways to child survival: results of a multi-country evaluation of Integrated Management of Childhood Illness*. *Health Policy and Planning*, 2005. **20**(supplement 1): p. i5-i17.
9. Bryce, J., et al., *The Multi-Country Evaluation of the Integrated Management of Childhood Illness Strategy: Lessons for the Evaluation of Public Health Interventions*. *American Journal of Public Health*, 2004. **94**: p. 406-415.
10. Grijalva, C.G., et al., *Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis*. *The Lancet*, 2007. **369**: p. 1179-1186.
11. Victora, C.G., et al., *Measuring impact in the Millennium Development Goal era and beyond: a new approach to large-scale effectiveness evaluations*. *Lancet*, 2010. **July 9. 2010**(DOI: 10.1016/S0140-6736(10)60810-0).
12. Deaton, A., *Randomization in the tropics, and the search for the elusive keys to economic development*. 2009, Princeton University.
13. Jahn, A., et al., *Population-level effect of HIV on adult mortality and early evidence of reversal after introduction of antiretroviral therapy in Malawi*. *The Lancet*, 2008. **371**: p. 1603-1611.
14. Gow, J.A., *The adequacy of policy responses to the treatment needs of South Africans living with HIV (1999-2008): a case study*. *Journal of the International AIDS Society*, 2010. **12**(37): p. doi:10.1186/1758-2652-12-37.