Aiming High and Falling Low: the SADA-Northern Ghana Millennium Village Project

Edoardo Masset, Jorge García Hombrados and Arnab Acharya[*]

December 2, 2019

Abstract

This article assesses the impact of the Northern Ghana Millennium Village Project. We estimate project effects on the Millennium Development Goals (MDGs) indicators using a difference-in-difference approach applied to matched villages and households using a sub-classification of the propensity score. The project improved some MDG indicators but, with few exceptions, impacts were small and core welfare indicators, such as monetary poverty, undernutrition and child mortality, remained unaffected. We found no spillover effects of the project to neighbouring areas and no displacements of development expenditure by local government and NGOs. We assessed the cost-effectiveness of the intervention and concluded that MVP did not produce the expected cost-saving synergies. We attribute the lack of impact to poor project design, redundancy of the interventions, and excessively high expectations. **_Keywords:_** Millennium Villages Project, integrated rural development, Millennium Development Goals. **_JEL Codes:_** O12, O13, O15.

[*]**Edoardo Masset**: Corresponding author. CEDIL at the London School of Hygiene and Tropical Medicine. Email: Edoardo.Masset@lshtm.ac.uk. LIDC, 20 Bloomsbury Square, London, UK. Postcode: WC1A 2NS. **Jorge Garcia Hombrados**: London School of Economics. Email: j.garcia-hombrados@lse.ac.uk. Department of Social Policy, LSE. 2nd Floor, Old Building, Houghton St, London. Postcode:wc2a 2ae. **Arnab Acharya**: Honorary Senior Lecturer at the London School of Hygiene and Tropical Medicine. Email: Arnab.Acharya@lshtm.ac.uk.

Acknowledgements

1 Introduction

The SADA-Northern Ghana Millennium Village Project[1] (MVP) was one of 15 Millennium Village projects implemented in Sub-Saharan Africa since 2006 by the Millennium Promise, the Earth Institute at Columbia University and the United Nations Development Program. The project was designed following the recommendations of the United Nations Millennium Project (UN Millennium Villages Project, 2005) to show that the Millennium Development Goals (MDGs) could be achieved in rural Africa at a small geographic scale and at relatively small cost through interventions in multiple sectors (Sanchez et al., 2007).

Evaluating the impact of MVP was important for two reasons. First, the MVP was arguably one of the most influential and debated development projects ever implemented. The project emerged from the Millennium Summit held in New York on 6-8 September 2000, in which 147 heads of state adopted the Millennium Declaration, which committed countries to achieve the MDGs by 2015. The project was endorsed by the UN Secretary General, and by numerous prime ministers, bilateral donors, philanthropists, academics, and celebrities. It was for 10 years the UN flagship anti-poverty programme. We estimated the cost of the project in 15 localities to be at least $300 million.[2] Given the resources involved, there were compelling accountability reasons for evaluating the intervention.

A second reason for evaluating the MVP was that the project revisited a scarcely tested development theory. The theory underlying the MVP was that a 'big push' of simultaneous interventions in multiple sectors would generate complementarities and cost-saving synergies that would lift people permanently out of poverty. There is a large literature testing the presence of poverty traps and people's and countries' ability to overcoming them (Kraay and McKenzie, 2014). However, rarely an intervention set out the goal of 'breaking the poverty

---

[1]The Savannah Accelerated Development Authority (SADA) was an agency of the Government of Ghana established in 2010 with the goal of coordinating development efforts in the northern regions of Ghana. SADA was involved in the design and implementation of the MVP and the project came to be known as the SADA-Northern Ghana Millennium Village Project. SADA discontinued operations and ceased to exist in 2016.

[2]The figure is based on the average cost of $120 per person over 5 years for 500,000 beneficiaries reported on the Millennium Village Project website (http://millenniumvillages.org/the-villages/).

trap', as in the case of the MVP, thus providing evaluators the opportunity to test the theory empirically.

The evaluation of MVP posed two main challenges. First, since the project was implemented in a cluster of contiguous villages, randomisation was not possible. Second, the project implementers opposed an evaluation design comprising multiple arms that would have allowed the identification of the impacts of single interventions as well as their interactions. In our study, we addressed the first challenge by using a difference-in-difference design between matched project and control villages. We addressed the second challenge by estimating the impact of the intervention as a 'package' rather than trying to disentangle the effects of single interventions, and we used cost-effectiveness analysis to infer the presence of cost-saving synergies.

Our study finds that the impact of MVP on the MDGs was limited, and that core welfare indicators such as monetary poverty, child mortality and under-nutrition were not affected. Two placebo tests confirmed the plausibility of the results, and additional analysis ruled out spill-over effects or displacements of development expenditure. Our cost-effectiveness analysis suggests that rather than generating cost-saving synergies, the project was more expensive than the sum of the costs of the components separately implemented, a result in agreement with the historical record on the success of integrated development projects (OED, 1988).

Our results differ from the positive results found by evaluations of MVP conducted by researchers affiliated with the Earth Institute. These studies did not include either a control group or a baseline and their results are difficult to interpret. Remans et al. (2011) found a reduction in stunting and improvements in food security, child care and infectious diseases in MVP villages. However, their study was a before-after analysis of changes in project areas in comparison to national trends. Pronyk et al. (2012) reported improvements in skilled birth attendance, bed-net use, malaria incidence, access to sanitation, and child mortality, but some of these results were subsequently retracted (Bump et al., 2012; Pronyk, 2012).

In addition, the study did not include a baseline, and impact estimations relied on data collected from villages 'similar' to project villages at the end of the project (Pronyk and Palm, 2009). Mitchell et al. (2018) found statistically significant impacts on 30 of 40 welfare outcomes. Their results, however, overstate the impact of the intervention as some of the outcomes considered were not MDG indicators, whilst others represented the same welfare outcome in different ways.[3] The main limitation of the study was that it was retrospective and the estimations of project effects were single differences between project and control observations at the end of the project. For all these studies, it is difficult to discern the impact of the intervention from underlying trends.

That impacts of MVP should be estimated after subtracting underlying trends was eloquently shown by Clemens and Demombynes (2011), who illustrated the risk of bias present in simple before-after comparisons, absent a valid control group. They compared trends in skilled birth attendance, ownership of mobile phones, stunting and primary school attendance, in three Millennium Villages, to trends in rural areas of the same countries using Demographic and Health Survey data (DHS). However, they did not take account of pre-existing differences between the people in the project and comparator villages; and not for all comparisons did they find data at matching times. Wanjala and Muradian (2013) found large impacts of MVP in the village of Sauri in Kenya on maize yields and on consumption of own produced food. Their analysis was limited to agricultural outcomes, it estimated a single difference at the end of the project, used a very small sample (nine project localities and five control localities with a total of 411 observations), and employed matching methods using variables that could have been affected by the project. Michelson and Tully (2018) used data on land transactions in the area surrounding the Sauri village in Kenya between

---

[3]Some of the outcomes considered by Mitchell et al. were project outputs rather than outcomes or MDG indicators, for example farmers' use of fertiliser and farmers' use of seeds, while other indicators were highly correlated, for example they used four separate indicators for bednet use (bednet use by children under-5, bednet use by pregnant women, bednet ownership and bednet correct use), four indicators for anaemia incidence (under-5 anaemia, school-aged anaemia, women anaemia, and men anaemia) and four indicators for malaria incidence (under-5 malaria, school-aged malaria, women malaria and men malaria). The claim of improving 30 of 40 outcomes, though correct, overstates the overall impact.

1999 and 2013 and found that the project did not affect land prices suggesting that project benefits were small or not incorporated into land prices.

Our study differs from all previous attempts to evaluate the impact of MVP in various ways. Firstly, our evaluation design was more rigorous. This is the only prospective evaluation of MVP that collected baseline data, and four follow-ups survey rounds, from project and control villages using large samples of panel observations. Secondly, we built the best possible control group by selecting matched control villages neighbouring the project villages, and by carrying out further matching of observations at the household level using state-of-the-art methodology. Thirdly, our study measured the largest number of MDG indicators, including hard-to-measure outcomes that figure prominently in the MDGs, such as monetary poverty, child mortality, undernutrition, and malaria incidence. Fourthly, this is the only study estimating spill-over effects of the intervention through a stratification of the control group by distance. This is important because it has been argued that benefits of the MVP could extend beyond the project villages, and that impact of MVP would be hard to detect due to the potential contamination of observations in the control group. Lastly, this is the only study attempting a cost-effectiveness analysis of MVP. This is relevant because positive impacts from an intervention of such scale were expected, but doubts had been raised regarding its economic efficiency. Finally, unlike some of the previous evaluations, our study was fully independent and underwent several rounds of scrutiny by external reviewers. The evaluation was registered with the Register for International Development Impact Evaluations (RIDIE) hosted by 3ie (Masset, 2015), and the analysis presented in this paper closely followed the pre-analysis plan (Masset, 2014).

As noted, despite some positive impacts, we found mostly null results, suggesting that the intervention was ineffective. From this we draw two main lessons. The first is that implementing several interventions in different sectors at the same time, contrary to expectations, does not automatically ensure the emergence of complementarities and cost-saving synergies. Our study does not provide evidence in support of 'big push' small-area interventions. The

second lesson is that untested interventions should always be piloted and evaluated. Resources could have been saved, and a more effective programme could have been designed, if an evaluation of MVP had been built in its early stages of implementation.

The paper is structured in the following way. The next section describes MVP activities and the theory underpinning the intervention. Section 3 describes the evaluation design, the data, and the methods used for the estimation of project effects. Section 4 presents the impact of the intervention on the MDG indicators. Section 5 examines the robustness of the results to different placebo tests. Section 6 investigates displacement and spill-over effects. Section 7 assesses the cost-effectiveness of MVP, and section 8 concludes.

2   The SADA-Northern Ghana Millennium Village

The Northern Ghana MVP was implemented between May 2012 and December 2016, with funding of £11 million from the UK Department for International Development (DFID). It provided a package of services in agriculture, health, education and infrastructure to a cluster of 35 villages with an approximate population of 3,900 households and 27,000 individuals. The project sites were identified using poverty maps (Coulombe and Wodon, 2007a,b), and spanned between the West Mamprusi and Builsa districts.

MVP did not have a theory of change articulating how the activities would produce outputs and outcomes. Instead, project managers had at their disposal a series of tables mapping potential interventions to outcomes for the following sectors: agriculture, community mobilisation, education, energy, environment, gender, infectious diseases, maternal and child health, health services, nutrition, and water. The interventions were to be adapted to the local context, and detailed studies were carried out prior to project start, in order to assess the specific needs of the area. The activities had to be identified and implemented with the collaboration of local authorities and communities, requiring the project to build local capacity and to coordinate with local institutions. Some activities were trialled and dis-

continued if unsuccessful, while others were adopted through a process of learning-by-doing. The implementation of the project required enormous administrative efforts and struggled to coordinate activities with local governments and communities. These implementation difficulties are similar to those that have been deemed responsible for the failure of similar integrated rural development projects in the past (OED, 1988; Clemens and Demombynes, 2011; Masset, 2018).

The project spent an average $88 per capita per year distributed in the following way across sectors: health (39.5%), agriculture (18.0%), infrastructure (18%), education (15.4%), environment (1.3%), and community mobilisation (7.8%). Information on costs was not available at a level of detail sufficient to assess the relevance of each project activity. Inspection of cost data suggested that investments were substantial in the following activities: deployment of community health workers (CHWs), provision of fertiliser, teacher hiring, and construction of roads, clinics and schools. More specifically, MPV carried out the following activities by sector.

In agriculture the project started by forming and training cooperatives in each community. In the first year, the project donated seeds and fertiliser for the production of maize, soy and rice. In subsequent years, inputs were provided through subsidised loans. Ten tractors were made available for hire to farmers at a heavily subsidised price. Eight agricultural extension officers were recruited to complement the 14 already employed by the government. Extension officers operated through more than 150 'lead farmers', who provided training on agronomic practices to fellow farmers through demonstration plots and farm visits. The project rehabilitated, or built anew, warehouses for the storage of agricultural output. Contacts with traders and local markets were facilitated in order to give farmers access to better prices.

In the education sector the project started its activities with community sensitisation. Community education workers were hired and trained to hold community meetings and to support parents and teachers associations. School infrastructure was improved. Four schools

were rehabilitated and other eight were newly built. Schools were provided with toilet facilities and playgrounds, sporting equipment, teaching materials, books and computers. Pupils were periodically donated stationery and other school supplies. Community education workers conducted a real-time monitoring of pupils' reading and numerical skills using mobile technology to assign remedial education. New teachers were hired and trained. Teacher's quarters were built and salary top-ups were provided. Communities and parents were sensitised on the benefits of girls' schooling. Separate toilets were built for girls. Girls were given free sanitary pads and a limited number of secondary school scholarships.

In the health sector the project started with 'quick wins': the distribution of mosquito bed-nets, immunisations, and de-worming. The cornerstone of MVP in the sector was the deployment of community health workers. More than fifty CHWs were hired, who would visits all household monthly and perform malaria tests, provide vitamin A, de-worming tablets, condoms, and counselling on nutrition and health. An e-Health system was established that enabled real-time patient tracking and referral through mobile phones. Health facilities were rehabilitated or newly built and furbished with medical equipment, nutritional supplements, drugs, vaccines, and testing kits. Two ambulances were procured and drivers were recruited. Health facilities were staffed with nurses, physicians' assistants, midwives and laboratory technicians. The project paid for community health workers salaries and provided top-ups for Ghanaian Health Service staff. A comprehensive ante-natal care package was provided through periodic monitoring of pregnant women by CHWs, screening and prevention at health clinics, tetanus toxoid immunisation, and micronutrient supplementation. The project promoted deliveries attended by trained midwives at health facilities. All individuals were registered with the National Health Insurance Service in the first year and registration was promoted in the following years. Several outreach initiatives disseminated information on infectious diseases, HIV prevention, family planning, breastfeeding and nutrition.

Infrastructural interventions, in addition to the construction of health and education facilities, focused on roads, water, and sanitation. Eleven roads were rehabilitated, nearly

80km of new roads were completed, and 70 culverts were constructed. The project implemented a Community-Led Total Sanitation programme in all communities, which resulted in the construction of pit latrines for nearly every household. The project awarded contracts for the construction of 23 boreholes and rehabilitated 49 water points. MVP lobbied the Ministry of Energy to expand services in the area and secured an agreement to extend the national grid to 32 of the 35 MVP communities. Remaining gaps in access to energy were addressed through the installation of solar panels at the community level and the provision of energy-saving cook stoves. The project also lobbied telecom operators to expand mobile data connectivity in the area.

The Millennium Villages were established as an experimental application of the recommendations made by the Millennium Project to end African poverty (UN Millennium Villages Project, 2005). Two distinct but related ideas underpinned the MVP: the classical theory of poverty trap, and the presence of output complementarities in implementing small area interventions. We will discuss these two theories in turn.

Poverty trap is one of the oldest theories in development economics and early formulations can be found in the writings of Rosenstein-Rodan (1943), Nurske (1953), Leibenstein (1957), Myrdal (1957), and Jorgenson (1961). More recent expositions can be found in Murphy et al (1989), and Bowles et al. (2006). Generally speaking, a poverty trap is a self-reinforcing mechanism that causes poverty to persist. In the canonical threshold poverty trap model (Azariadis and Stachurski, 2005; Carter and Barrett, 2006; Ghatak, 2015; Ravallion, 2016), poverty traps are generated by feedback mechanisms whereby current country capital (A_t), or household assets, is a function of capital in the previous period (A_{t-1}). The dynamic growth path produced by this iterative process is illustrated in Figure 1. The 45 degrees line represents all points in which capital in the current period and capital in the previous period are equal. Any point other than C brings capital either to point A or B. Point A is a poverty trap: small capital increases beyond A lead capital to converge again to point A. A fundamental implication of this model is that capital increases beyond the threshold

(point C) will push a country, or household, on a sustainable growth path towards the high equilibrium point B. The poverty trap model provides therefore the theoretical foundations for 'big push' policy prescriptions.

Figure 1: The classical poverty trap model
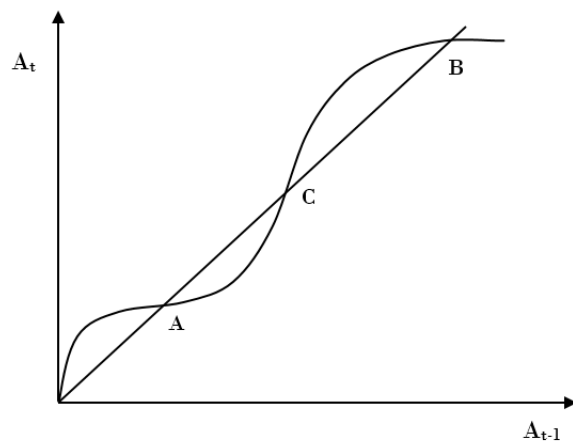
Sachs et al. (2004), in the study that set the background to the design of MVP, illustrated three different versions of the poverty trap model of Figure 1. In the first, poor human and physical infrastructure result in very low returns to investments. In the second, poor people's inability to save results in an inability to accumulate capital. In the third, people are trapped in poverty by rapid demographic growth. They identified five factors that trapped sub-Saharan Africa in poverty: high transport costs and small market size, poor agricultural productivity, burden of malaria and poor health, a history of colonisation and depredation, and slow technology diffusion from abroad. They consequently proposed a 'big push' solution to the African poverty trap through interventions in agricultural productivity; health, nutrition and family planning; primary education; urban infrastructure and services; science and technology; gender equity; and regional integration.

The MVP was conceived as a "proof of concept" demonstration project through which the African poverty trap could be overcome, and the MDGs achieved, at village-scale (Sanchez et al., 2007). The success of the MVP would provide the evidence required for scaling

up the same interventions at the national level thus leading to the end of poverty in the African continent. The MVP was not meant to address all potential sources of poverty traps, particularly those arising at the macro level, and it focused on micro-level poverty traps resulting from market imperfections, such as missing credit markets, and poor human and physical factor endowments. Nevertheless it maintained the underlying hypothesis, typical of poverty trap models, that by significantly augmenting the capital stock of households and communities beyond a threshold level, poverty could be permanently escaped .

Despite considerable theoretical development, poverty traps have been difficult to track empirically and there is conflicting evidence on whether they exist or not, either at the country or household level (Kraay and McKenzie, 2014). One attractive feature of evaluating MVP was therefore that it offered an indirect way of testing poverty trap theory. The project, if implemented correctly, was expected to permanently reduce poverty over several welfare dimensions as captured by the MDG indicators.

The second idea underlying MVP was that the project would unlock synergies by exploiting complementarities between interventions (Sachs et al., 2004; Sanchez et al., 2007). To understand this point, we may think of different project activities as inputs to the production of an outcome. For example, the provision of fertiliser increases income directly, but this impact is augmented by education interventions, which increase farm productivity. It is further augmented by health interventions, which reduce working days lost, and by road improvements, which favourably affect input and output prices. Supposing that MVP produced outputs whose impact could be reduced to a single welfare measure, such as income or disability-adjusted life years, then we could express the effect of complementarities through introducing a little notation. Let x_i be the inputs to produce output y_i and $\bar{x}_i$ be a fixed level of inputs producing level $\bar{y}_i$ if used in a stand-alone production process. Then integrated projects are efficient if the jointly used inputs produce greater welfare as in:

$$Y_{joint}(\bar{x}_1, \ldots, \bar{x}_n) > \sum_{i=1}^{n} \bar{y}_i(\bar{x}_i) \tag{2.1}$$

whereby the joint production of the welfare outcome (Y_{joint}) by employing $\bar{x}_1, \ldots, \bar{x}_n$ inputs (project activities) is larger than the sum of the outputs obtained by implementing the activities separately.

Equation 2.1 is not useful for practical analysis. This is because firstly, MVP interventions operated in several dimensions of welfare and it would be difficult to construct a composite indicator capturing impacts operating at different levels. For example, years of educational gains do not easily translate to income, at least in the short run, and increased access to toilet facilities does not translate easily to health-related quality of life indicators. Building a single welfare measure would require a large number of assumptions. Secondly, it would be difficult to identify what inputs were used for any single specific purpose in the joint production, unless we were conducting a multi-arm experiment, in which the interventions were implemented solely in some communities and jointly in others.

A more feasible way to estimate gains from joint production is to measure synergies at the cost level; if inequality 2.1 holds then the following inequality must also hold:

$$C_{joint}(\bar{y}_1, \ldots, \bar{y}_n) < \sum_{i=1}^{n} c_i \bar{y}_i \tag{2.2}$$

where C_{joint} is the joint cost of producing $\bar{y}_i \ldots \bar{y}_n$ outputs, and c_i stands as the unit costs of each type of output when produced separately. Outputs are produced at the same level on both sides of the inequality. The overall cost of MVP should be smaller (right-hand-side of the relation) than the sum of the costs of achieving the same outputs in different sectors implementing separate interventions.

The advantage of using inequality 2.2 rather than 2.1 is twofold. First, synergies generate cost savings, and to assess the efficiency of joint production there is no need to aggregate welfare out of different outputs in a single metric. Second, both sides of 2.2 are observable; the joint cost of observed outputs from MVP can be compared to the aggregation of unit costs multiplied by the same individual levels of outputs produced through stand-alone production processes. In the absence of results from a multi-arm experiment, we can compare the cost

of joint production from MVP, which is known, to costs for obtaining the same level outputs in similar settings, as they are reported in the literature.

3   Study design, data, and empirical methods

To evaluate MVP we employed a difference-in-difference (DiD) design using matched project and control villages, and matched project and control households. We selected the control sites by matching them to project villages using village-level data. After collecting household data, we matched project households to control households using household-level data.

Our project sample consists of all 35 project villages. We selected the control villages using a one-to-one matching based on village-level characteristics obtained from the 2000 and 2010 population censuses, supplemented by observations collected in the field (Masset et al., 2013). We paired every project village to a control village from each of two strata of equal size. One stratum was composed of villages in the vicinity of the project and the other stratum was composed of villages far from the project. Hence, the sample of control villages includes 34 villages in the vicinity of the intervention, and 34 villages far away from the intervention but within the district boundaries. We oversampled control communities with the primary goal of estimating spill-over effects and with the secondary goal of allowing further matching at the household level.[4]

The baseline survey aimed at collecting data from 755 project households and 1,496 control households based on power calculations to detect impacts of acceptable size for key outcomes (Masset et al., 2013).[5] Baseline surveys were conducted between April and September 2012. Follow-up rounds were conducted every year for the following four years (2013 to 2016), and at the same time of the year. The diagram in Figure 2 illustrates the structure of the sample at the selection, follow-up and analysis stage.

---

[4]Further details of the stratification and a map of project and control villages are reported in the technical appendix.

[5]Ex-ante power calculations are reported in the published version of the evaluation design (Masset et al., 2013), and can be found in the technical appendix.

Figure 2: Flow diagram of project and control households
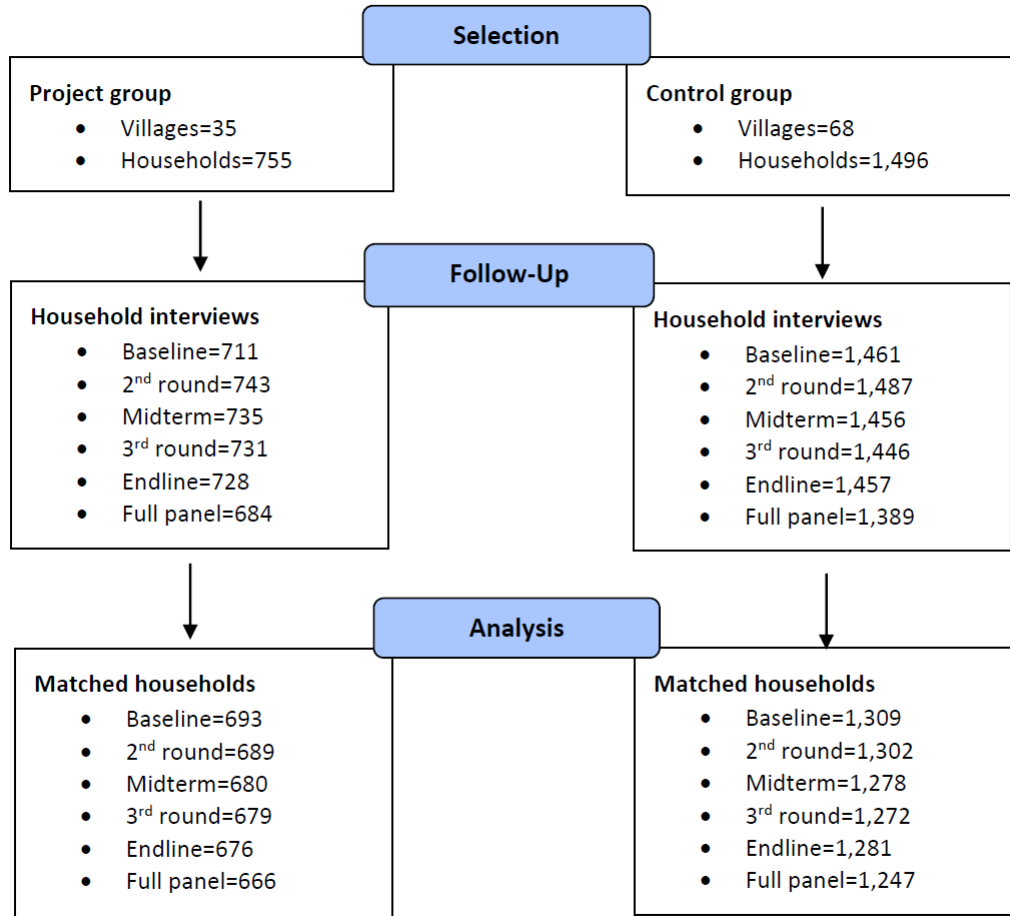
The survey instruments consisted of: a household questionnaire modelled on the Living Standard Measurement Surveys; an adult questionnaire modelled on the DHS questionnaires; an anthropometric questionnaire collecting height and weight of children under-5; a malaria module recording blood specimens of children under-5 collected through finger/heel poke; and a community questionnaire collecting information on village characteristics. Table 1 shows the observations collected at each survey round by each survey tool. We assessed the quality of the data and concluded that it was comparable to the quality of data normally collected by the Ghanaian Statistical Office and other agencies.[6]

---

[6]We assessed the quality of the data in various ways. We inspected income and expenditure data using an application of the *Benford's law* of first digits (Judge and Schechter, 2009). First digits of reported income and expenditure figures were compared to a Benford distribution to detect anomalies or data fabrication.

Table 1: Survey instruments used in the MVP evaluation

	2012	2013	2014	2015	2016
Household questionnaire	2,172	2,230	2,191	2,177	2,185
Adult female questionnaire (15-49)	2,837	-	3,241	-	2,837
Adult male questionnaire (15-49)	1,628	-	1,835	-	1,671
Anthropometry (children under-5)	1,933	-	1,670	-	1,513
Blood tests (children under-5)	805	-	1,121	-	968
Community questionnaire	103	103	103	103	103

Table 2 reports the number of households in the project and control groups across survey rounds and related attrition rates. Attrition rates were very small and even at the endline did not exceed 5%. More importantly, attrition rates were very similar in the project and control group. We analysed the characteristics of attriters using logistic regression and we could not find structural differences between the project and control samples.

Table 2: Household attrition in project and control areas

Sample	2012	2013	2014	2015	2016
Households panel (project)	711	707	697	689	684
attrition		0.6%	2%	3.1%	3.8%
Household panel (control)	1461	1,454	1,424	1,391	1,389
attrition		0.5%	2.5%	4.8%	4.9%
All households	2172	2,161	2,121	2,080	2,073
attrition		0.5%	2.3%	4.2%	4.6%

We estimated impacts using a DiD analysis of observations matched on a sub-classification of a propensity score. The sub-classification approach to matching is a state-of-the-art matching approach to estimating causal effects of non-randomly allocated interventions, that has not been widely used, and yet has a number of advantages in comparison to traditional

---

We assessed the quality of child mortality data using the method of overlapping cohorts (Hill, 2013), which postulates that mortality rates in a given year should be the same when calculated using data from different survey rounds. Finally, we assessed the quality of data on height, weight and hemoglobin by comparing sample standard deviations to values commonly found in other datasets.

matching approaches (Imbens and Rubin, 2015). An early empirical application of this approach, similar to to the one employed here, can be found in Dehejia and Whaba (1999). We summarise here our estimation strategy, which is explained in greater detail in the technical appendix. First we divided the sample in sub-classes of project and control observations with similar values of the estimated propensity score. We then estimated causal effects of the intervention within each sub-class as if the assignment of the intervention had been random within each sub-class. Finally we calculated the average treatment effect of the intervention as the weighted average of the sub-classes effects, the weights being given by sub-class sizes. A key element of this approach is the identification of the optimum number of sub-classes, for which we employed the following algorithm. We considered adequate a sub-class when a t-test on the difference in the propensity score between the project and the control observations in that sub-class was less than 1.69. If the null was rejected we split the the sub-class at the median value of the propensity score and we tested the difference in the propensity scores in the two newly formed sub-classes. We proceed iteratively in this way until we obtained sufficiently homogeneous sub-classes for which the null of no difference in the propensity score could not be rejected. The estimation of project effects within small samples does not compromise precision of the estimates and statistical power. An additional advantage of sub-classification is that, like stratification in randomised trials (Athey and Imbens, 2016; Imbens, 2011), by estimating differences within homogeneous sub-classes of observations, it reduces the variance of the estimates and therefore increases statistical power, provided the covariates employed for generating the sub-classes are good predictors of the outcomes.

Imbens and Rubin (2015) recommend combining sub-classification with covariance adjustment through linear regression. We therefore estimated project effects within each sub-class using OLS and adjusting for a set of covariates that were correlated with the outcomes. The combination of sub-classification and regression adjustment has a number of advantages. First it reduces the bias of the estimates, because sub-classification alone does not eliminate all the bias associated with differences in the covariates. Second, regression adjustment

improves the precision of the estimates and increases efficiency. Finally, since regressions are estimated within sub-classes that are homogeneous in the values of the covariates, the estimates are less sensitive to the specification of the regression function. We calculated standard errors of the average treatment effect using 500 bootstrap replications at the cluster level. At each replication we drew a different sample, we ran the regressions at sub-class level, and we calculated a different average treatment effect and its standard error.

In the regressions, we estimated the impact of the intervention, as the average of year-specific impacts with respect to the baseline.[7] The DiD model over five time periods ($t = 0, \ldots, 4$) with individual-level variables, such as nutritional status or school attendance, is:

$$y_{it} = \alpha + \sum_{t=1}^{4} \beta_t T_{it} + \gamma P_i + \delta P_i POST_i + \sum_{j=1}^{n} \zeta_j X_{ji0} + \epsilon_{it} \tag{3.1}$$

where y_{it} is the outcome observed for the observation i at time $t = 0, \ldots, 4$, where 0 is the baseline and $t = 1, \ldots, 4$ are four follow-up rounds. T_t are dummy variables for each follow-up, P is a dummy variable equal to 1 if the observation is in the project group. $POST$ is a dummy variable equal to one for every observation collected after the baseline. The coefficient δ of the interaction of the project variable with the $POST$ variable is the average DiD effect. The X_{j0} are baseline values of ($j = 1, \ldots, n$) covariates that improve the balance between the project and the comparison group.

With household-level variables, such as poverty and access to water, we used fixed-effects models at the household level to account for the impact of time-invariant unobservables.[8] The average project effects with panel data were estimated using:

$$y_{it} = \alpha_i + \sum_{t=1}^{4} \beta_t T_{it} + \sum_{j=1}^{n} \gamma_j X_{jit} + \delta P_i POST_i + \epsilon_{it} \tag{3.2}$$

Average project effects were estimated, as before, by the δ coefficients. The covariates (X_j)

---

[7]In the technical appendix we present the result of all year-specific impact estimates.

[8]With individual-level variables, such as prevalence of undernutrition and child mortality, fixed effects model can be used only drastically reducing the sample because most children under-5 at the baseline were no longer in the sample at the endline.

are time-varying variables that are not affected by the project and include weather shocks (floods and droughts) and household demographic composition.[9]

The propensity score was estimated with a logit model including a fixed set of covariates that we believed were highly correlated with the outcomes (household size, age and education of head of household, land size, and wealth), whilst other covariates, their squares and interactions, were added stepwise, based on their statistical significance. After estimation, the sample was trimmed to remove observations with extreme values of the propensity score and the trimmed sample was used to re-estimate the propensity score and to build new sub-classes of the propensity score.

We assessed the validity of our matching algorithm by checking covariate balance at baseline with three tests. The first test assessed the global balance of each covariate across sub-classes. Several covariates were out of balance before matching (first column of Table 3), but after matching, the Z-values of the tests were very low (second column of Table 3). In the second test we assessed the balance of each covariate within all sub-classes. To do this we ran regressions of the covariates on sub-class dummies and the interactions of sub-class and project dummies. We found only two covariates out of balance (third column of Table 3). In the third test, we assessed balance within each sub-class for each covariate, producing 136 different tests. We plotted the Z-values of the tests using a quantile-quantile plot and we found that most observations lied above or along the 45° line suggesting that the results were very similar to those obtained from random draws of Z-values from a normal distribution.[10]

We conclude this section with a brief note on addressing the multiple hypothesis testing problem. Since we were estimating a large number of project effects there was a good chance

---

[9]In four cases we had to use inverse probability weights (IPW) rather than regression analysis, because the indicator could not be calculated within regression models. Child and infant mortality rates were estimated using the synthetic cohort probability method used in DHS surveys (Masset, 2016). The gender parity indicator is a population-level ratio of the number of girls in primary school over boys over small age-groups. Similarly, the consumption share of the poorest quintile is the population-level ratio of expenditure by the poorest 20% over total expenditure. In the IPW approach, biases arising from differences in the covariates are removed by weighting observations by the inverse of the propensity score. The weights (w) for the estimation of average treatment effects are calculated as: $w = \frac{P}{e} + \frac{(1-P)}{(1-e)}$, where e is the propensity score and P is equal to 1 if the observation is in the project group and 0 otherwise.

[10]See Figure 6 of the technical appendix.

Table 3: Test of covariance balance

Covariate	Unadjusted T-test	Z-value test across strata	F-value test within strata
Household size	-0.62	-0.13	0.71
Age of head	-1.17	0.25	1.42
Education of head	-1.49	-0.24	1.14
Cultivated land	-2.29	-0.31	0.47
Wealth	-1.49	0.01	0.69
Remittances	-3.91	-0.03	2.19
Millet farm	-3.22	-0.10	0.75
Rice farm	3.01	0.09	0.43
Drought shocks	3.37	-0.07	3.75
Flood shock	-2.67	-0.03	1.40
Isolated household	-3.24	-0.03	0.44
Months food insecure	2.79	0.12	0.44
Farmer household	2.58	0.27	1.07
Bank access	-2.84	-0.06	1.02
Metal roof	0.53	-0.09	0.52
Distance to drinking water	-1.13	0.00	0.59
Groundnut farm	-1.88	0.46	1.26

of finding effects when there was none, simply as the result of chance. For example, with 29 different hypotheses about impacts on the MDG indicators, and with a statistical significance threshold of 10%, there is a 95% probability of finding at least one statistical significant effect even if the intervention has absolutely no impact (the probability is $1 - (1 - \alpha)^N$, where α is the level of statistical significance and N is the number of hypotheses). To address this problem we employed the False Discovery Rate (FSR) of Benjamini-Hochberg (Efron and Hastie, 2016), which calculates the proportion of false discoveries among the rejected hypotheses. We ordered the p-values of each test in ascending order and indexed them by $i = 1, \ldots, N$ and we rejected all the null hypotheses whose p-value was less than $\frac{i}{N}q$. This test is more conservative than standard statistical tests and it rejects fewer null hypotheses, thus reducing false positives.

4 Impact of MVP on the MDGs

MVP increased people's participation in the activities promoted by the intervention. Table 4 shows participation rates in some of the services offered by the project. The project increased participation in agriculture (membership of farmer groups, agricultural training, access to loans, fertiliser use), education (access to school meals), and health (visits to clinics, visits by community health workers, family planning and post-natal care). With the exception of access to iron tablets, impacts were large and statistically significant. We regressed participation on household and individual characteristics and we found that it was not correlated to education, wealth, land or other relevant characteristics. The project provided services without targeting specific groups, and no specific group took advantage of the intervention.

MVP had the goal of breaking the poverty trap by achieving the MDGs (Sachs et al., 2004; Sanchez et al., 2007). The MDGs were eight welfare goals relating to poverty, education, gender equality, child mortality, maternal mortality, morbidity, environmental sustainability and global cooperation, that had been adopted in 2000 by 191 countries and 22 international organisations under the aegis of the United Nations. Each goal had targets (18 in total) to be met by 2015, and a total of 48 measurable indicators. The MDGs were not adopted without criticism (see for example the articles in Black and White (2004), but became very popular among donors and development agencies, before being replaced by the Sustainable Development Goals in 2016. A tremendous data collection exercise enabled us to estimate impacts on 29 of the 39 MDG household-level indicators.[11]

We start our analysis by testing baseline differences between project and control areas (Table 5). Only three MDG indicators were different at baseline: the parity ratio in primary school, the use of mosquito bed-nets for children under-5, and the prevalence of ante-natal care. Other indicators were fairly similar, suggesting that the two groups were comparable.

---

[11]A full description of all MDG indicators and of how they were calculated using our survey data can be found in the technical Appendix

Table 4: Impact of MVP on project participation

	Baseline CV	Baseline diff MV	Impact midterm	Impact endline
Any household member received a loan	3.3	1.6	10.6***	-0.1
		(1.448)	(2.097)	(0.479)
Used fertiliser	39.4	-7.7	18.1***	10.8**
		(5.039)	(4.151)	(5.307)
Child had a school meal on previous day	20.1	8.9	16.3**	12.1
		(7.188)	(7.239)	(8.044)
Member of Farmer Group	7.1	0.5	11.1***	8.5***
		(1.479)	(1.730)	(1.482)
Training in agriculture	12.7	3.3	31.4***	7.1**
		(2.613)	(2.588)	(3.109)
Visited by CHW	36.8	-10.4***	28.9***	47.6***
		(3.572)	(5.009)	(2.940)
Went to health clinic	32.7	-3.7	14.1***	8.0***
		(2.612)	(2.405)	(2.799)
Family planning at facility	20.4	-4.3*	12.9***	7.1**
		(2.227)	(4.108)	(2.948)
Iron tablets	73.9	-10.2		-5.9*
		(7.540)		(3.195)
Child weighted after birth	23.7	-5.1		16.7***
		(3.916)		(5.436)

Note: cluster-level standard errors in parentheses. Three stars (***) is statistical significance at 1%, two stars is 5%, and one star is 10%.

Also of interest is the comparison between the study area and the rest of the country (last column of Table 5). The data suggest that the area was much poorer in monetary terms than the rest of Ghana. Child mortality rates were higher, maternal health and health services were poorer, and access to public facilities such as mobile technology and sanitation was also poorer. However, for some indicators, such as school attendance, undernutrition, gender parity, malaria prevalence and access to water, the area compared relatively well with the rest of the country.

Table 5: Baseline MDG indicators in project and control areas

MDG	Baseline CV	Baseline diff MV	P-value	Ghana
Proportion of population below $1.25 (PPP) per day	83.52	0.00	0.963	
Proportion of population below the national poverty line	88.1	-0.5	0.836	24.2[1]
Poverty gap ratio	48.7	1.0	0.699	7.8[1]
Consumption share of poorest quintile	7.2	-0.6	0.777	5.4[2]
Employment to population ratio	79.5	-3.2	0.217	72.5[2]
Proportion of employed people living below $1.25 (PPP) per day	33.5	-3.1	0.515	
Proportion of own account and contributing family workers in total employment	95.9	-3.6	0.080	71.0[2]
Underweight prevalence (children under-5)	16.4	-1.8	0.345	11.0[3]
Proportion of population below the food poverty line	66.5	0.9	0.798	8.4[1]
Net attendance ratio in primary education	69.6	-9.3	0.056	70.6[3]
Completion rate in primary education	74.5	2.9	0.602	78.9[2]
Young adults (15-24) literacy rate	32.9	-2.5	0.505	85.1[3]
Ratio of girls to boys in primary education	1.0	0.2*	0.006	1.0[3]
Share of women employed in the non-agricultural sector	37.6	-7.4	0.431	
Under-5 mortality rate	10.4	-3.5	0.030	6.0[3]
Infant mortality rate	7.0	-2.6	0.097	4.1[3]
Measles immunisation rate (children under-2)	50.9	10.4	0.015	89.3[3]
Proportion of births attended by skilled health personnel	30	-0.7	0.902	73.7[3]
Contraceptive prevalence rate	9.8	-0.3	0.884	26.7[3]
Adolescent birth rate	14.6	10.5	0.059	7.6[3]
Ante-natal care coverage	84.1	6.4*	0.009	97.0[3]
Proportion of young adults (15-24) with correct HIV knowledge	8.7	1.1	0.528	23.6[3]
Proportion of children under-5 sleeping under insecticide treated bed-nets	54.6	-23.8*	0.000	46.6[3]
Malaria prevalence (children under-5)	22.5	-2.9	0.471	26.7[3]
Proportion of children under-5 with fever treated with anti-malarial	36.3	-1.3	0.863	48.5[3]
Proportion of the population using an improved drinking water source	72.5	-0.1	0.946	64.2[3]
Proportion of the population using an improved sanitation facility	8.9	0.3	0.923	15.0[3]
Fixed telephone subscriptions rate	0.2	-0.2	0.194	1.1[1]
Mobile telephone usage rate	45.7	3.2	0.419	80.3[1]

Note: P-values based on cluster standard errors. A star (*) represents a statistically significant coefficient at 10% with respect to a critical value adjusted by FDR. Data sources:[1]GLSS(2012-13),[2]World Bank WDI (2012),[3]DHS (2014).

The impact of the intervention on the MDG indicators is illustrated in Figure 3. The figure plots point estimates together with confidence intervals adjusted by the critical value of the FDR. We coloured in red those MDG indicators that represent project outputs rather than welfare outcomes. As mentioned, the programme did not have a theory of change describing how activities would affect outputs, and how these in turn would affect welfare outcomes. The construction of such theory, given the wide range of activities implemented and their interrelations, proved to be difficult. However, the simplest possible theory of change would predict the interventions to accomplish their own outputs, as delineated in the

project logic models. Some of these outputs overlapped with the MDGs indicators, namely immunisation rates, births attended by skilled professionals, access to modern contraception, ante-natal care, safe water, safe sanitation, use of bed-nets, and of anti-malarial drugs. It should be expected that the delivery of, for example, mosquito bed-nets and antenatal care, should increase their reported use. However, theory would be more ambiguous regarding whether those outputs would lead to the desired welfare outcomes, such as a reduction in malaria incidence and infant mortality. The distinction is therefore useful for a first understanding as to whether the lack of project impact was the result of poor implementation or poor design.

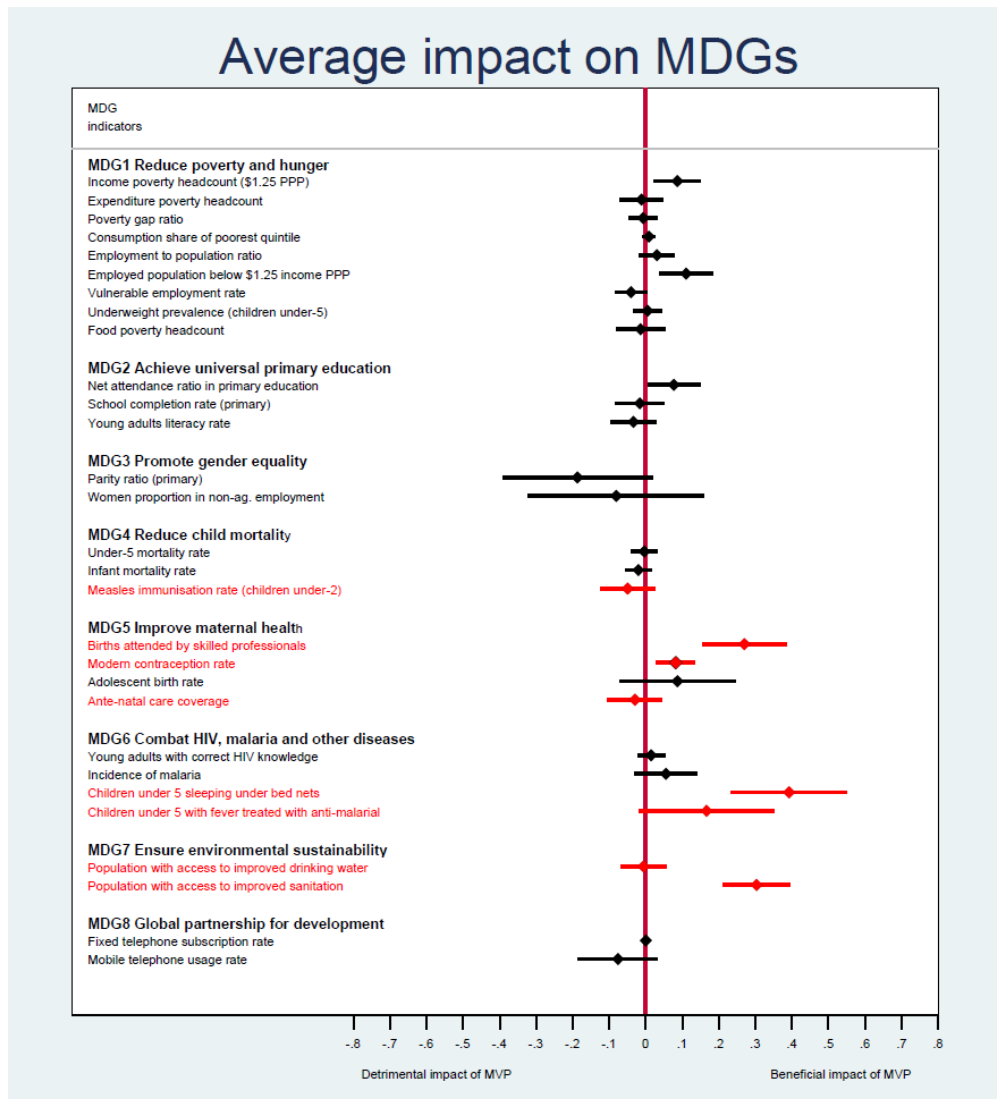Figure 3: Average impact of MVP on MDG indicators

Table 6: Impact of MVP on MDG indicators

MDG indicator	Simple Diff-in-diff	Regression adjustment	Kernel matching	Coarsened matching	Sub-classification matching
Proportion of population below $1.25 (PPP) per day	-11.58*	-9.01*	-10.03*	-6.68	-8.66*
	(2.74)	(2.87)	(3.37)	(3.10)	(2.73)
Proportion of population below the national poverty line	1.88	3.09	1.32	-1.75	1.16
	(3.20)	(3.35)	(2.99)	(4.31)	(3.41)
Poverty gap ratio	0.50	1.32	-0.83	-0.62	0.71
	(2.29)	(2.24)	(2.25)	(2.97)	(2.32)
Consumption share of poorest quintile	1.29*	1.30	1.68	2.59	0.87
	(0.50)	(0.62)	(0.87)	(1.50)	(0.86)
Employment to population ratio	2.93	3.16	1.50	1.01	3.06
	(2.48)	(2.42)	(2.39)	(2.93)	(2.41)
Proportion of employed people living below $1 (PPP) per day	-10.42*	-10.92*	-10.56*	-7.45	-11.04*
	(3.13)	(3.13)	(3.38)	(3.30)	(3.03)
Proportion of own account and contributing family workers in total employment	4.75*	4.84*	4.63*	5.53*	4.02
	(1.68)	(1.68)	(1.91)	(1.61)	(1.93)
Underweight prevalence (children under-5)	-1.04	-0.66	1.26	3.04	-0.52
	(2.30)	(2.36)	(2.32)	(3.10)	(2.28)
Proportion of population below the food poverty line	1.19	2.42	-0.38	-0.50	1.42
	(3.79)	(3.77)	(4.32)	(5.03)	(3.84)
Net attendance ratio in primary education	7.64*	7.14*	8.90*	10.87*	7.69*
	(2.89)	(2.79)	(3.09)	(3.91)	(3.16)
Completion rate in primary education	-1.34	-1.67	-0.29	-5.24	-1.62
	(2.88)	(2.94)	(3.59)	(3.50)	(3.79)
Young adults (15-24) literacy rate	-4.94	-4.13	2.09	-4.79	-3.36
	(2.88)	(2.83)	(3.23)	(4.39)	(3.33)
Ratio of girls to boys in primary education	-15.96	-17.51	-24.10*	-26.12	-18.70
	(8.45)	(8.60)	(8.46)	(12.70)	(9.68)
Share of women employed in the non-agricultural sector	-16.08	-16.12	-21.33	-22.52	-8.06
	(10.09)	(9.26)	(12.64)	(21.52)	(13.40)
Under-5 mortality rate	0.58	-1.90	1.37	2.82	0.41
	(2.03)	(3.28)	(2.00)	(2.10)	(2.07)
Infant mortality rate	2.18	1.40	2.17	2.65	2.02
	(1.85)	(1.91)	(2.01)	(2.00)	(1.89)
Measles immunisation rate (children under-2)	-1.97	-2.89	-2.25	-2.66	-4.95
	(4.44)	(3.68)	(4.36)	(4.83)	(3.70)
Proportion of births attended by skilled health personnel	28.82*	26.59*	29.35*	21.79*	27.00*
	(4.76)	(4.45)	(5.03)	(6.96)	(4.50)
Contraceptive prevalence rate	7.68*	7.70*	7.62*	9.14*	8.22*
	(1.68)	(1.64)	(1.98)	(2.56)	(2.13)
Adolescent birth rate	-4.79	-4.65	-4.30	-13.46	-8.67
	(4.75)	(4.09)	(4.61)	(8.04)	(8.31)
Ante-natal care coverage	-3.68	-3.84	-3.66	-1.86	-2.94
	(3.53)	(3.50)	(4.12)	(4.47)	(4.06)
Proportion of young adults (15-24) with correct HIV knowledge	0.49	0.20	1.50	-0.50	1.47
	(1.98)	(1.96)	(2.00)	(3.02)	(2.05)
Malaria prevalence (children under-5)	-0.15	-1.20	-4.22	-5.65	-5.53
	(4.24)	(4.20)	(4.54)	(6.39)	(4.28)
Proportion of children under-5 sleeping under insecticide treated bed-nets	38.96*	39.19*	35.21*	45.30*	39.24*
	(5.38)	(5.37)	(5.41)	(5.98)	(5.85)
Proportion of children under-5 with fever treated with anti-malarial	23.34*	22.32*	17.88	10.26	16.65
	(7.60)	(7.80)	(8.49)	(11.39)	(8.52)
Proportion of the population using an improved drinking water source	-2.42	-2.16	-2.07	-1.47	-0.61
	(3.24)	(3.26)	(2.41)	(3.27)	(3.83)
Proportion of the population using an improved sanitation facility	29.92*	29.69*	31.18*	27.53*	30.33*
	(2.91)	(2.81)	(3.62)	(2.88)	(3.16)
Fixed telephone subscriptions rate	0.03	0.03	0.13	0.50	0.04
	(0.12)	(0.11)	(0.16)	(0.48)	(0.12)
Mobile telephone usage rate	-9.05	-8.95	-8.47	-12.83	-7.60
	(5.11)	(5.20)	(5.12)	(6.06)	(5.23)

Note: Coefficients and standard erros were multiplied by 100 to improve readibility as percent differences. Standard errors were calculated using 500 bootstrap replications. A star (*) represents a statistically significant coefficient at 10% with respect to a critical value adjusted by FDR.

All MDG indicators are binary, and the DiD impacts have a simple interpretation as

percent points differences between the project and the control group. In some cases, a beneficial impact of the intervention is a reduction in the indicator (for example a reduction in poverty), while in other cases a beneficial impact is an increase (for example an increase in births attended by skilled professionals). To simplify the reading of the chart we changed the sign of the impacts in such a way that beneficial impacts are all on the right-hand-side of the chart and detrimental impacts are on the left-hand-side. The vertical line represents no impact.

The size of the effects and their standard errors are reported in the last column of Table 6.[12] The table also reports effects estimated using four other specifications: simple DiD without covariate adjustment, DiD with regression adjustment, DiD with kernel matching, and DiD with coarsened exact matching. These results are presented as a robustness check of the sensitivity of the results to alternative specifications. The results of the three matching specifications are in agreement both in terms of effect sizes and statistical significance and do not differ substantially from the effects estimated using the unmatched sample with and without covariates.[13]

The project had a beneficial impact on seven indicators (one fourth of total). With two exceptions (use of mosquito bed-nets and access to toilet facilities) impacts were small, the differences being under 10 per cent points. Many of the beneficial impacts occurred on output indicators, while few and small impacts were observed on outcome indicators. There were no detrimental effects, with the possible exception of a reduction of the gender parity ratio

---

[12]Year-specific impacts of the intervention are reported in the technical Appendix.

[13]The effects reported in the Kernel matching column were obtained using DiD regressions with Kernel weights. The weights were obtained by matching observations using the set of 20 baseline covariates that were found to be associated with MV at a minimum 15% of statistical significance (see Appendix 4 on matching methods). Kernel matching was performed using the psmatch2 Stata programme of Leuven and Sianesi (2003). The results reported in the Coarsened matching column of Table 6 were obtained using DiD regressions with weigths calculated using coarsened exact matching. The variables used in coarsening included all basic variables that were found to be associated with MVP at a minimum 15% of statistical significance and all other variables whose normalised difference at the baseline between the project and the control group was larger than 0.15 (household size, age of head of household, education of head, land size, total household wealth, number of months of food insecurity, isolated household and household having a working migrant). The weights were calculated using the CEM Stata programme of Blackwell et al. (2009).

in primary education. We now discuss all impacts in more detail.

MVP had a very limited impact on *eradicating extreme poverty and hunger* (goal 1). Interestingly, the project reduced income poverty by nearly 9 percent points but did not reduce expenditure poverty or food poverty. The intervention did not reduce inequality as measured by the poverty gap and by the consumption share of the bottom quintile. Finally, the project did not reduce the prevalence of underweight children, an indicator that combines acute and chronic malnutrition. Possibly as a result of agricultural interventions, the project increased total employment and employment of family workers on own account, though these effects were not statistically significant.

The project had a moderate impact on *achieving universal primary education* (goal 2). Primary school attendance increased by 7.7 percent points in comparison to control areas. This impact was remarkable considering that attendance rates in primary school were already high. The project, however, did not improve completion rates, thus failing to promote retention among pupils already in school. Finally, the intervention did not improve adult literacy rates.

MVP did not affect goal 3 (*improving gender equality and empowering women*). Indeed, the ratio of girls to boys in primary school decreased, though the effect was not statistically significant in most specifications. The project brought to school more boys than girls, possibly because more girls were already in school at baseline. The project did not increase the proportion of women engaged in non-agricultural activities such as micro-enterprises.

MVP did not reduce child mortality (goal 4). We found no impact either on infant or under-5 mortality. Actually the impact of MVP was negative, meaning that mortality decreased more rapidly in control areas. The project did not improve measles vaccination rates either.

MVP made significant progress towards *improving maternal health* (goal 5). The proportion of births attended by a skilled professional improved by 27 percent points. The proportion of women using modern contraceptive methods also improved by 8 percent points. The

rate of teenage pregnancies decreased, but the effect was not statistically significant. Finally, the project did not improve the coverage of ante-natal care.

The project had some effects on *combating HIV, malaria and other diseases* (goal 6). The proportion of children sleeping under mosquito bed-nets improved by 39 percent points. Prevalence of malaria among children decreased, though the effect was not statistically significant. There was a large impact on the proportion of children with fever treated with anti-malarial drugs (nearly 17 percent points), though the effect was not statistically significant. The proportion of adults with basic knowledge of HIV did not improve and remained worryingly low.

The project had a mixed impact on goal 7 (*ensure environmental sustainability*). It did not improve access to safe drinking water, a rather disappointing result given the considerable investments made in the construction of boreholes. On the other hand, access to improved toilet facilities improved by 30 percent points.

Finally, the project did not affect goal 8 (*developing a global partnership for development*). Ownership of telephone landlines did not increase and remained totally absent in the area. More surprising was the lack of impact on the use of mobile phones considering the efforts made by the project to improve access.

In sum, the project achieved, at best, mixed results. The count of statistically significant impacts is low. Even excluding some indicators that were not explicitly targeted by the intervention, such as adult literacy and landline connections, the project only improved about 25% of the indicators. More importantly, core welfare indicators (poverty, nutrition and mortality) were totally unaffected.

5 Sensitivity analysis

We assessed the robustness of our results with two placebo tests. In the first test we estimated the impact of the intervention on outcomes that were not affected by the intervention

(pseudo-outcomes approach). In the second test we estimated the impact of an intervention that did not take place, using only the sample of control observations (pseudo-intervention approach).

In the pseudo-outcomes approach we used lagged values of variables targeted by the intervention. Our surveys included retrospective questions on school attendance, cultivated land, livestock holdings, agricultural output, wages, and micro-enterprise revenues. We were not able to use the data on agricultural and business incomes because the quality was poor, whilst the sample of observations on wages was too small. School attendance was reported for one year before the baseline while in all other cases we collected data for two years before the baseline. We estimated impacts before and after the baseline (Table 7). After the baseline, the project had a positive impact on all three variables, and the effects were statistically significant in the case of school attendance and livestock holdings. Before the baseline, we found no impact on the same variables. This result can be interpreted as inability to reject the hypotheses of parallel trends, thus reassuring our DiD strategy.

Table 7: Impact of MVP on pseudo-outcomes

	Pseudo-effect t-1	Pseudo-effect t-2	Project effect 4-year average
Net attendance rate (primary)	0.012		0.077**
	(0.015)		(0.032)
Cultivated land (ln of acres)	0.002	0.029	0.092
	(0.026)	(0.034)	(0.062)
Livestock holdings (ln of value in Cedis)	0.128	0.032	0.333***
	(0.193)	(0.134)	(0.123)

Note: cluster-level standard errors in parentheses. Three stars (***) is statistical significance at 1%, two stars is 5%, and one star is 10%.

In the pseudo-intervention approach we split the control group by administrative region, after removing from the sample the stratum of 'nearby' villages that could have been affected by the project via spill-over effects. Households of the districts of Builsa and West

Mamprusi belong to different ethnic groups and are differently administered. Of all possible ways of splitting the control group, the administrative split is very conservative for testing purposes because we would expect to find some differences in outcomes in the absence of any intervention, simply because baseline characteristics, and possibly trends, in the two districts were different. We estimated the impact of a 'Builsa' intervention on 29 MDG indicators. We found only two statistically significant effects: the proportion of children sleeping under bed-nets, and household access to sanitation (see Table 8).[14]

---

[14]We also estimated the impact of the intervention separately in the two districts to see whether the intervention was more effective in one of the districts and we found no differences.

Table 8: Impact of pseudo-intervention on the MDGs

MDG	DiD (Control Far Builsa vs Control Far West Mamprusi)
Proportion of population below $1.25 (PPP) per day	0.006
	(0.919)
Proportion of population below the national poverty line	0.019
	(0.621)
Poverty gap ratio	-0.024
	(0.559)
Consumption share of poorest quintile	0.014
	(0.152)
Employment to population ratio	-0.047
	(0.222)
Proportion of employed people living below $1.25 (PPP) per day	0.004
	(0.956)
Proportion of own account and contributing family workers in total employment	-0.016
	(0.100)
Underweight prevalence (children under-5)	0.088
	(0.086)
Proportion of population below the food poverty line	0.010
	(0.878)
Net attendance ratio in primary education	0.063
	(0.197)
Completion rate in primary education	-0.035
	(0.390)
Young adults (15-24) literacy rate	0.024
	(0.704)
Ratio of girls to boys in primary education	-0.034
	(0.782)
Share of women employed in the non-agricultural sector	0.105
	(0.649)
Under-5 mortality rate	0.053
	(0.201)
Infant mortality rate	0.037
	(0.274)
Measles immunisation rate (children under-2)	-0.172
	(0.016)
Proportion of births attended by skilled health personnel	0.086
	(0.241)
Contraceptive prevalence rate	-0.025
	(0.502)
Adolescent birth rate	-0.055
	(0.545)
Ante-natal care coverage	0.028
	(0.600)
Proportion of young adults (15-24) with correct HIV knowledge	-0.007
	(0.853)
Malaria prevalence (children under-5)	-0.148
	(0.142)
Proportion of children under-5 sleeping under insecticide treated bed-nets	0.358*
	(0.000)
Proportion of children under-5 with fever treated with anti-malarial drugs	0.066
	(0.655)
Proportion of the population using an improved drinking water source	0.184
	(0.011)
Proportion of the population using an improved sanitation facility	0.107*
	(0.002)
Fixed telephone subscription rate	0.000
	(0.992)
Mobile telephone usage rate	0.014
	(0.837)

Note: P-values based on cluster-level standard errors in parentheses. A star (*) represents a statistically significant coefficient at 10% with respect to a critical value adjusted by FDR.

6 Displacements and spill-over effects

The impact of MVP could have been dampened by displacement and spill-over effects. Displacement effects occur when the intervention produces geographical reallocations of expenditures by the government or NGOs. Spill-over effects occur when households in control areas have access to the services offered by the intervention.

Interventions do not occur in a vacuum. Governments and NGOs implement interventions in project and control areas, some of which are similar to project interventions. DiD assumes that there are no differential changes in government and NGOs expenditures after an intervention, which might be unrealistic for large scale interventions such as MVP. The direction of bias produced by displacements is difficult to predict. In its early stages of implementation, it was suggested that MVP could encourage the implementation of similar interventions in the control areas by a process of diffusion. For example, MVP could pioneer the use of a technology that was then adopted by other agencies in the control areas. MVP project managers also claimed that MVP would act as a catalyst for projects run by other agencies, by leveraging private investments and government projects. It was also observed that, given the scale of its operations, MVP could displace projects implemented by other agencies. For example, NGOs and government agencies operating in the project area may decide to move operations to the control areas after the adoption of MVP.

We investigated these effects using community-level data on projects implemented by the government, the local district assembly, and by NGOs. Data collected from community leaders reported an average of 1.8 projects per village at baseline in areas such as: construction of boreholes, schools, teaching quarters, clinics, roads and bridges; distribution of books and school uniforms; electrification; improved latrines; cash transfers; school feeding; solar panels and rehabilitation of dams. There were no differences in the number of projects in project and control villages at the baseline (Table 9). We estimated the impact of MVP on the number of projects in the control areas. We found that after the intervention there was on average one more NGO project in MVP areas, which likely reflects the implementation of

33

MVP. MVP had a positive impact on projects implemented in control areas, but this impact was very small and not statistically significant. This means that we can rule out the hypothesis that the project siphoned out resources from the control areas, and that displacement effects, if any, were minimal.

Table 9: Impact of MVP on project in control areas

	NGO projects Coeff.	Government projets Coeff.	District assembly projects Coeff.
Baseline difference	-0.011	-0.079	0.162
	(0.199)	(0.162)	(0.268)
Project impact in control areas	-1.084***	0.124	0.118
	(0.114)	(0.181)	(0.300)

Note: Cluster-level standard errors in parentheses. Three stars (***) is statistical significance at 1%, two stars is 5%, and one star is 10%.

Next we considered spill-over effects. If project benefits extend to control areas, the comparisons are biased and the effects underestimated. For example, mothers and children from control areas might have access to the health services provided by the intervention. To test the presence of spill-over effects we used the stratification of the control group by distance. We first compared near and far control communities, and we did it separately for the districts of Builsa and West Mamprusi. The two districts are inhabited by two different ethnic groups with different structures of social relations. In addition, communities in West Mamprusi are more geographically dispersed and remotely located. There are therefore reasons to expect different spill-over effects in the two districts.

We first assessed the impact of MVP on project participation (Table 10). We found only one positive, but statistically insignificant, impact consistent with spillover effect: visits to health clinics. The other estimated impacts were either of a sign inconsistent with spill-over effects, or too small to produce an impact on the MDG indicators.

Finally, we estimated spill-over effects on the MDG indicators (Table 11). We could not find a coefficient that was significantly different from zero. Some coefficients had signs in

Table 10: Participation spillovers

	DD impact in near CV	DD impact in near CV West Mamprusi	DD impact in near CV Builsa
Participation in farmers' group	0.012	-0.002	0.028
	(-1.072)	(-0.152)	(-1.541)
Participation in agricultural training	-0.003	-0.112***	0.128***
	(-0.220)	(-4.769)	(-4.511)
Visited a health facility	0.032	0.018	0.050
	(-1.52)	(-0.636)	(-1.622)
Visited by a CHW	-0.050**	-0.032	-0.079**
	(-2.395)	(-1.156)	(-2.530)
Sample size	9,296	5,189	4,107

Note: coefficients are average marginal effects from logit model including cross-derivatives. Z-values in parentheses based on cluster standard errors. Three stars (***) is statistical significance at 1%, two stars is 5%, and one star is 10%.

accord with our expectations. For example, births attended by skilled professionals increased in nearby communities while child mortality rates decreased. However, other coefficients had unexpected signs. For example, prevalence of malaria increased in nearby communities, while ante-natal care decreased. Overall, these results do not support the hypothesis that a wave of benefits spread from project to control areas.

Table 11: Spillover effects

MDG	DiD (Near CV vs Far CV)	DiD (Builsa: Near CV vs Far CV)	DiD (West Mamprusi: Near CV vs Far CV
Proportion of population below $1.25 (PPP) per day	-0.060	-0.100	-0.026
	(0.180)	(0.022)	(0.732)
Proportion of population below the national poverty line	-0.022	-0.032	-0.014
	(0.471)	(0.528)	(0.752)
Poverty gap ratio	0.003	-0.027	0.033
	(0.913)	(0.516)	(0.460)
Consumption share of poorest quintile	-0.020	-0.054	-0.014
	(0.019)	(0.002)	(0.121)
Employment to population ratio	0.014	0.016	0.010
	(0.693)	(0.629)	(0.818)
Proportion of employed people living below $1 (PPP) per day	-0.047	-0.081	-0.025
	(0.371)	(0.073)	(0.747)
Proportion of own account and contributing family workers in total employment	0.042	0.055	0.031
	(0.015)	(0.074)	(0.074)
Underweight prevalence (children under-5)	0.002	-0.045	0.010
	(0.953)	(0.270)	(0.850)
Proportion of population below the food poverty line	-0.023	-0.085	0.036
	(0.641)	(0.156)	(0.628)
Net attendance ratio in primary education	0.057	-0.010	0.090
	(0.077)	(0.848)	(0.032)
Completion rate in primary educationn	0.083	0.094	0.051
	(0.006)	(0.130)	(0.340)
Young adult literacy rate	-0.055	-0.089	-0.014
	(0.229)	(0.209)	(0.852)
Ratio of girls to boys in primary education	-0.138	-0.049	-0.196
	(0.121)	(0.675)	(0.109)
Share of women employed in the non-agricultural sector	-0.043	-0.041	-0.032
	(0.848)	(0.618)	(0.944)
Under-5 mortality rate	-0.025	-0.026	-0.029
	(0.328)	(0.334)	(0.345)
Infant mortality rate	-0.025	-0.007	-0.039
	(0.205)	(0.479)	(0.050)
Measles immunisation rate (children under-2)	0.006	0.076	-0.021
	(0.924)	(0.408)	(0.776)
Proportion of births attended by skilled health personnel	0.125	0.147	0.067
	(0.010)	(0.110)	(0.223)
Contraceptive prevalence rate	0.006	-0.021	0.010
	(0.782)	(0.584)	(0.714)
Adolescent birth rate	0.015	0.052	-0.014
	(0.826)	(0.649)	(0.891)
Ante-natal care coverage	-0.034	-0.124	0.036
	(0.459)	(0.028)	(0.588)
Proportion of young adults (15-24) with correct HIV knowledge	0.048	0.030	0.066
	(0.042)	(0.551)	(0.013)
Malaria prevalence (children under-5)	0.109	0.252	-0.031
	(0.148)	(0.013)	(0.729)
Proportion of children under-5 sleeping under insecticide treated bed-nets	-0.068	-0.242	0.023
	(0.387)	(0.381)	(0.807)
Proportion of children under-5 with fever treated with anti-malarial	0.157	0.198	0.139
	(0.141)	(0.192)	(0.327)
Proportion of the population using an improved drinking water source	-0.009	-0.117	0.069
	(0.911)	(0.057)	(0.480)
Proportion of the population using an improved sanitation facility	-0.058	-0.067	-0.053
	(0.139)	(0.026)	(0.331)
Fixed telephone subscription rate	-0.001	0.001	No change
	(0.774)	(0.864)	
Mobile telephone usage rate	0.008	-0.168	0.127
	(0.883)	(0.025)	(0.059)

Note: P-values based on cluster standard errors. A star (*) represents a statistically significant coefficient at 10% with respect to a critical value adjusted by FDR.

7   Cost-effectiveness

We categorised cost data under six sectors: health, education, agriculture, infrastructure, community development and environment, but we were not able to identify the costs of specific activities, for example, malaria prevention. The total cost of MVP was \$16.2 million, of which \$4.7 million was spent on management, administration and operation, monitoring and evaluation, technical assistance and research. The average yearly per-capita cost was \$123. However, after removing administrative costs and after annuitising the costs of infrastructure expenditures - many of which were conducted towards the end of the project and delivering benefits in the future - the total cost of MVP was \$10.2 million corresponding to \$88 per capita per year.[15] As a comparison, the cost of the ultra-poverty graduation programme in Northern Ghana - another multi-sectoral programme implemented with the goal of breaking the poverty trap - was \$75 per capita (Banerjee et al., 2015). Health was the largest project area, followed by infrastructure (Table 12). Community development was implemented across sectors and environmental interventions were negligible.

Table 12: MVP cost per-capita by sector

	Health	Education	Agriculture	Environment	Infrastructure	Community Dev	Total	Yearly Costs
Cost[1]	156	61	71	5	71	31	395	88
% of total	39.5	15.4	18.0	1.3	18.0	7.8		

Note: [1]These figures include cost of management, operating and administration, technical assistance and monitoring and evaluation.

Our ambitions to assess the cost-effectiveness of MVP were dwarfed by the lack of impacts. The calculation of internal rates of return and of cost-effectiveness ratios requires effect sizes that are larger than zero. We were nevertheless able to calculate cost-effectiveness ratios for those MDG indicators that were positively affected by the interventions: income, primary

---

[15]See the technical appendix for details of cost calculations.

school enrolment, access to sanitation, malaria prevalence, proportion of births attended by skilled professionals, and use of female contraception. As discussed in Section 2, we would expect MVP to be more cost-effective than other programmes in delivering these outcomes because of cost-saving synergies.

To assess the cost-savings hypothesis, we estimated the cost of achieving the same MDG indicator incurred by other projects independently implemented. To this aim, we undertook an extensive literature search to identify costs of similar interventions in similar contexts. The results are presented in Table 13.

Table 13: Comparison of cost of MVP with alternative interventions

MDG Indicator	Population-level impact	Cost using alternative program	Most likely MVP cost
Income poverty	$141, per capita, Total gain in 4.5 years: US$ 3.76 Million	US$ 2.82 Million	Lower Bound: US$1.82 Million, Upper Bound US$3.64 Million
Primary school attendance	2,153 School years	US$ 200,000	US$ 1.6 Million
Health outcomes: Malaria prevalence	891 fewer cases of malaria	All health costs:	US$ 4.2 Million
Increased healthcare usage, Skilled birth attendance Vaccination Contraception usage	Increased contact with CWHs by 48,500 832 all types of vaccination 2,112 year protection	US$ 400,000 to US$ 600,000	
Access to sanitation	8,187 people per year	Not Costed	Lower Bound: US$ 0.376 million
Approximate Total Costs		US$ 4 to US$ 5 million	US$9.5 to 10.2 million

Total income gains of MVP amounted to $141 per capita. We estimated the MVP cost to achieve this outcome between $1.8 and $3.6 million. In comparison, the poverty graduation programme (Banerjee et al., 2015), at a benefit cost ratio of 133%, would have produced the same return at a cost of $2.8 million. In education, MVP yielded 2,153 additional years of schooling. We estimated that this was achieved at the cost of $655 per year of schooling, which was much higher than the range of $5 to $91 to achieve the same outcome found in other studies (Baird et al., 2011; Evans and Popova, 2014). The project averted 891

cases of malaria, engendered 832 vaccinations, and produced 2,112 years of women using contraceptive methods. We estimated that the cost of malaria prevention (White et al., 2011), vaccination[16], and CHWs visits required to achieve the same outcomes, would have been $300-400 thousand (Doherty. and Govender, 2004), far lower than the $4.1 million spent by MVP.

We conclude that alternative projects could have achieved the same outcomes at a cost of $4 to 5 million, which is far below the $10.2 million spent by MVP. It appears that MVP, far from generating cost savings, delivered outcomes at a higher cost than comparable projects in similar contexts.

These results needs to be treated with caution because the comparator projects where not identical to MVP and they were implemented with populations of different characteristics, which could respond to the MVP interventions in different ways. Unlike the comparator projects considered in this section, MVP was designed to yield long-run benefits through system-wide interventions. To account for this we used costing methods adjusting for expenditures whose benefits run for several years and beyond the life of the project. We acknowledge that comparisons in this section are imperfect and only suggestive. The comparator costs come from programmes that had specific aims and sometime set in ideal small-scale environments; some even lacked proper rigour in calculating economic costs or ignored the cost of using current infrastructure. However, the large gap between the estimated costs of alternative ways of producing the same outcomes and MVP costs is revealing. We conclude that MVP was not cost effective because it did not yield a large enough impact when we observed a positive change and also the types of positive changes were only a few.

---

[16]see UNICEF 2016 https://www.path.org/publications/files/TS_opt_in_country_transport_rpt.pdf

8   Conclusions

In this paper we presented the results of the only prospective impact evaluation of the Millennium Village Project. The Northern Ghana MVP aimed at achieving the MDGs through an integrated package of investments in multiple sectors. Our evaluation found that the project affected some MDG indicators but the impacts were small and below expectations. Only about a quarter of the MDG indicators were beneficially affected and they mostly consisted of project outputs rather than welfare outcomes. Most sectors showed some beneficial results but no sector made definitive progress. The project did not seem to have unlocked cost-saving synergies. Our cost-effectiveness analysis, suggested that MVP was less cost-effective than similar interventions independently implemented.

The absence of impact was not the result of poor evaluation design. Our study built a valid control group and conducted the analysis using state of the art methodology for quasi-experimental designs. The study collected data from large samples of households and was sufficiently powered to detect the large impacts that the intervention had set out to achieve. The absence of impact was not the result of lack of people's involvement in the intervention either. Participation in the activities promoted by MVP was high and there is no evidence that segments of the population were either privileged or excluded by the project. The lack of impacts was not the result of project displacements or spill-over effects. Data on projects implemented by the government and NGOs suggest that displacement effects were minor, while our tests based on the geographic stratification of the control group suggest that no spill-over effects were present.

Having ruled out plausible alternative explanations we are left to speculate on the reasons for lack of impact. We identify three main reasons: poor project design, redundancy of the intervention, and excessively high expectations. Poor design and implementation is the first obvious explanation. The interventions implemented were not necessarily those mostly needed in this particular setting. For example, the project may have excessively focused on increasing school attendance or on reducing the prevalence of malaria in an

40

area where primary enrollment rates and malaria prevalence were similar to those prevailing in the rest of the country. There is also a long record of unsuccessful implementation of integrated development projects. The available literature on integrated rural development (OED, 1988; Clemens and Demombynes, 2011) reports that these projects often ran into difficulties deriving from the complexity of integrating multiple interventions across sectors. A qualitative institutional assessment of the project (Dogbe et al., 2018) found that MVP, similarly to integrated projects of the past, used a parallel structure to manage its activities with negative results in terms of efficiency and long term sustainability .

A second explanation is the partial redundancy of the interventions. Some MVP activities, for example the deployment of CHWs, the distribution of mosquito bed-nets, or the rehabilitation of schools, boreholes and clinics, were also implemented by the Ghanaian government and NGOs - though at a much smaller scale. To some extent the MVP may simply have come to substitute for the same interventions implemented by other agencies thus producing a limited additional impact. This does not mean that these MVP interventions were not efficacious, but that they did not bring much added value to ongoing trends.

A final explanation rests on the excessive expectations of project managers and designers. Rarely interventions set out to achieve such ambitious goals and we are not aware of any other evaluation based on the achievement of the MDGs. Surely all MVP interventions implemented produced some impact, but few produced the big improvements that the project designers had envisaged. The project expected to meet the MDGs over a period of five to ten years, a task that had been shown impossible to achieve (Clemens et al., 2007).

References

Athey, S. and Imbens, G. W. (2016). The eonometrics of randomised experiments. *arXiv.org*, Papers 1607.00698.

Azariadis, C. and Stachurski, J. (2005). *Poverty Traps*. North Holland, Dordrecht.

Baird, S., McIntosh, C., and Ozler, B. (2011). Cash or condition: Evidence from a randomized cash transfer program. *Quarterly Journal of Economics*, 126(4):1709–1753.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236):772–788.

Black, R. and White, H. (2004). *Targeting development: critical perspectives on the Millennium Development Goals*. Routledge, London and New York.

Blackwell, M., Iacus, S., King, G., and Porro, G. (2009). cem: Coarsened exact matching in stata. *The Stata Journal*, 9l(4):524–546.

Bowles, S., Durlauf, S., and Hoff, K. (2006). *Poverty Traps*. Princeton University Press, Princeton.

Bump, J. B., Clemens, M., Demombynes, G., and Haddad, L. (2012). Concerns about the millennium villages project report. *The Lancet*, 379(May 26):1945.

Carter, M. R. and Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *Journal of Development Studies*, 42(2):178–199.

Clemens, M. and Demombynes, G. (2011). When does rigorous impact evaluation make a difference? the case of the millennium villages. *Journal of Development Effectiveness*, 3(3):305–339.

Clemens, M., Kenny, C. J., and Moss, T. (2007). The trouble with the mdgs: Confronting expectations of aid and development success. *World Development*, 35(5):735–751.

Coulombe, H. and Wodon, Q. (2007a). Combining census and household survey data for better targeting: The west and central africa poverty mapping initiative. *MPRA paper*, No. 10483.

Coulombe, H. and Wodon, Q. (2007b). *Poverty, Livelihood, and Acess to Basic Services in Ghana*. The World Bank, Washington DC.

Dehejia, R. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94:1053–1062.

Dogbe, T., Apam, I., Salam, K., and Taufique, S. (2018). Institutional assessment. report.

Doherty., J. and Govender, R. (2004). The cost-effectiveness of primary care services in developing countries: A review of the international literature. *Disease Control Priority Project*, 37.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press, New York.

Evans, D. and Popova, A. (2014). Cost-effectiveness measurement in development: Accounting for local costs and noisy impacts. *World Bank Policy Research Working Paper*, 7027.

Ghatak, M. (2015). Theories of poverty traps and and anti-poverty policies. *World Bank Economic Review*, 29(Supplement 1):S77–S105.

Hill, K. (2013). *Direct estimation of child mortality from birth histories*. International Union for the Scientific Study of Population, Paris.

Imbens, G. (2011). Experimental designs for unit and cluster randomised trials. *Paper prepared for the International Initiative for Impact Evaluation, 3ie.*

Imbens, G. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences: an Introduction.* Cambridge University Press, New York.

Jorgenson, D. (1961). Development of the dual economy. *Economic Journal*, 71:309–334.

Judge, G. and Schechter, L. (2009). Detecting problems in survey data using benford's law. *Journal of Human Resources*, 4(1):1–24.

Kraay, A. and McKenzie, D. (2014). Do poverty traps exist? assessing the evidence. *Journal of Economic Perspectives*, 28(3):127–148.

Leibenstein, H. (1957). *Economic Backwardness and Economic Growth.* Wiley, New York.

Leuven, E. and Sianesi, B. (2003). Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html(version 4.0.11).

Masset, E. (2014). *Northern Ghana Millennium Village Impact Evaluation: Analysis Plan.* ITAD.

Masset, E. (2015). *Impact Evaluation of the SADA Northern Ghana Millennium Village Project.* RIDIE.

Masset, E. (2016). Syncmrates: Stata module to compute child mortality rates using synthetic cohort probabilities. *EconPapers* https://econpapers.repec.org/software/bocbocode/s458149.htm.

Masset, E. (2018). Integrated development, past and present. *IDS Bulletin*, 49(4):17–32.

Masset, E., Acharya, A., Barnett, C., and Dogbe, T. (2013). An impact evaluation design for the millennium villages project in northern ghana. *Journal of Development Effectiveness*, 5(2).

Michelson, H. and Tully, k. (2018). The millennium villages project and local land values: Using hedonic pricing methods to evaluate development projects. *World Development*, 101:377–387.

Mitchell, S., Gelman, A., Ross, R., Chen, J., Bari, S., Huynh, U. K., Harris, M. W., Sachs, S. E., Stuart, E. A., Feller, A., Makela, S., Zaslavsky, A. M., McClellan, L., Ohemeng-Dapaah, A., Namakula, P., Palm, C. A., and Sachs, J. D. (2018). The millennium villages project: a retrospective, observational, endline evaluation. *The Lancet*, 6:e500–e513.

Murphy, K., Shleifer, A., and Vishny, R. (1989). Industrialisation and the big push. *Journal of Political Economy*, 97(5):1003–1026.

Myrdal, G. (1957). *Economic Theory and Underdeveloped Regions*. Duckworth, London.

Nurske, R. (1953). *Problems of Capital Formation in Underdeveloped Countries*. Oxford University Press, Oxford.

OED (1988). *Desarrollo Rural: Experiencia del Banco Mundial*. The World Bank, Washington DC.

Pronyk, P. (2012). Errors in a paper on the millennium villages project. *The Lancet*, 379(May 26):1946.

Pronyk, P. M., Muniz, M., Nemser, B., Somers, M. A., McClellan, L., Palm, C. A., Huynh, U. K., Ben Amor, Y., Begashaw, B., McArthur, J. W., Niang, A., Sachs, S. E., Singh, P., Teklehaimanot, A., and Sachs, J. D. (2012). The effect of an integrated multisector model for achieving the millennium development goals and improving child survival in rural sub-saharan africa: a non-randomised controlled assessment. *Lancet*, 379(9832):2179–88.

Pronyk, P. M. and Palm, C. (2009). 09prt/8648: The millennium villages project: integrating the delivery of health and development interventions and assessing the impact on child survival in sub-saharan africa (nct01125618). *The Lancet*, http://www.thelancet.com/protocol-reviews/09PRT-8648.

Ravallion, M. (2016). *The Economics of Poverty: HIstory, Measurement and Policy*. Oxford University Press, New York.

Remans, R., Pronyk, P. M., Fanzo, J. C., Chen, J., Palm, C. A., Nemser, B., Muniz, M., Radunsky, A., Abay, A. H., Coulibaly, M., Mensah-Homiah, J., Wagah, M., An, X., Mwaura, C., Quintana, E., Somers, M. A., Sanchez, P. A., Sachs, S. E., McArthur, J. W., and Sachs, J. D. (2011). Multisector intervention to accelerate reductions in child stunting: an observational study from 9 sub-saharan african countries. *Am J Clin Nutr*, 94(6):1632–42.

Rosenstein-Rodan, P. (1943). Problems of industrialization of eastern and southeastern europe. *Economic Journal*, 53:202–211.

Sachs, J., Mcarthur, J., Shmidt-Traub, G., Kruk, M., Bahadur, C., Faye, M., and McCord, G. (2004). Ending africa's poverty trap. *Brookings Papers on Economic Activity*, 1(2004):117–240.

Sanchez, P., Palm, C., Sachs, J., Denning, G., Flor, R., Harawa, R., Jama, B., Kiflemariam, T., Konecky, B., Kozar, R., Lelerai, E., Malik, A., Modi, V., Mutuo, P., Niang, A., Okoth, H., Place, F., Sachs, S. E., Said, A., Siriri, D., Teklehaimanot, A., Wang, K., Wangila, J., and Zamba, C. (2007). The african millennium villages. *Proc Natl Acad Sci U S A*, 104(43):16775–80.

UN Millennium Villages Project (2005). Investing in development. a practical plan to achieve the millennium development goals. Journal article.

Wanjala, B. and Muradian, R. (2013). Can big push interventions take small-scale farmers out of poverty? insights from the sauri millennium village in kenya. *World Development*, 45:147–160.

White, M., Conteh, L., Cibulskis, R., and Ghani, R. (2011). Costs and cost-effectiveness of malaria control interventions, a systematic review. *Malaria Journal*, 10:337.