
NORTHERN GHANA MILLENNIUM VILLAGES IMPACT EVALUATION

ANALYSIS PLAN

Date: January 2015

By: Edoardo Masset

(with comments from Arnab Acharya and Patrick Nolen)

Submitted by Itad
In association with:



itad

Results in development



Acknowledgements

This report has been prepared by the team for the impact evaluation of the Millennium Villages Project. The team is composed of staff from Itad, the Institute of Development Studies, the London School of Hygiene and Tropical Medicine, and PDA-Ghana. The team is fully independent of the Earth Institute and the Millennium Promise. The principal author of this report is Dr Edoardo Masset, with additional comments from Dr Arnab Acharya and Dr Patrick Nolen. The report was edited by Dr Chris Barnett and proofread by Kelsy Nelson and Caitlin McCann. The team is very grateful to all the researchers and other colleagues that have assisted with the work of the evaluation. This Analysis Plan has been reviewed by staff at DFID, members of the Evaluation Advisory Group, and the independent Peer Review Group. All comments have been taken on board although the final text remains the full responsibility of the authors. Any views contained in this report do not necessarily represent those of DFID or of the people consulted.

Citation

Masset, E. 2014. Northern Ghana Millennium Villages Impact Evaluation: Analysis Plan. Itad, Hove.

Acronyms and Abbreviations

ANCOVA	Analysis of Covariance
ATE	Average Treatment Effects
ATT	Average Treatment Effects on the Treated
CPI	Consumer Price Index
CV	Control Village
DD	Difference-in-Difference
DHS	Demographic and Health Survey
EI	Earth Institute
GDP	Gross Domestic Product
GER	Gross Enrolment Rate
GLSS	Ghana Living Standards Survey
GPS	Global Positioning System
GSS	Ghana Statistical Service
HIV	Human Immunodeficiency Virus
IDD	Initial Design Document
ISDN	Integrated Services Digital Network
JHS	Junior High School
M&E	Monitoring & Evaluation
MDGs	Millennium Development Goals
MV	Millennium Village
MVP	Millennium Villages Project
NGO	Non-Governmental Organisation
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
PSM	Propensity Score Matching
PTA	Parent Teacher Association
WHO	World Health Organisation
WLL	Wireless Local Loop
VIP	Ventilated Improved Pit latrine
VOIP	Voice Over IP

Report

Analysis Plan for the Northern Ghana Millennium Villages Impact Evaluation

Table of Contents

ACKNOWLEDGEMENTS	2
ACRONYMS AND ABBREVIATIONS	3
EXECUTIVE SUMMARY	6
1. ACHIEVEMENT OF THE MDGS	8
2. IMPACT HETEROGENEITY	13
3. MULTIPLE OUTCOMES TESTING	14
4. PROGRAMME PARTICIPATION	15
5. ECONOMETRIC METHODS	17
5.1 DIFFERENCE-IN-DIFFERENCES.....	17
5.2 COMBINING DD AND MATCHING.....	19
5.3 THREATS TO VALIDITY.....	21
6. EXPLORATORY ANALYSIS	25
6.1 SPILLOVER EFFECTS.....	25
6.2 IMPACT ON NON-MDG OUTCOMES.....	26
6.3 THEORIES OF CHANGE AND QUALITATIVE WORK.....	28
6.4 RETURNS TO FACTORS.....	28
6.5 BREAKING THE POVERTY TRAP.....	29

TABLES

TABLE 1. THE MDG OUTCOME INDICATORS.....	8
TABLE 2. PARTICIPATION IN PROGRAMME ACTIVITIES AND EXAMPLES OF CHARACTERISTICS DETERMINING PARTICIPATION.....	16
TABLE 3. TREND DATA COLLECTED BY THE HOUSEHOLD AND COMMUNITY SURVEYS IN 2012.....	22
TABLE 4. NON-MDG OUTCOME INDICATORS.....	27

Disclaimer

This report has been prepared by Itad specifically for the services outlined below, and is in all respects subject to the negotiation, agreement and signature of a specific contract of engagement. The details included in this report shall not be disclosed to any other party, or be duplicated, used or disclosed in whole or in part for any purpose other than to evaluate the proposed service to be performed by Itad.

This confidentiality clause applies to all pages and information included in this report.

Executive summary

This is the analysis plan for the impact evaluation for the Millennium Villages Project in Northern Ghana. The purpose of the plan is to set out more specific details about how the team intends to analyse the quantitative datasets to estimate the impact of the intervention in 2016/17. This document builds on the earlier Initial Design Document (IDD),¹ which set out the overarching design and the methodology for data collection and analysis. The impact analysis is divided into a *confirmatory* component, directed to assess the achievement of the Millennium Development Goal (MDG) targets using hypotheses testing and methods of causal inference, and an *exploratory* component, directed to assess impact on non-MDG targets and on the causes for success and failure in achieving some of the targets using a wider range of methods of causal inference and directed to formulate hypotheses that could be rigorously tested in the future. We summarise here the main elements of the plan, as follows:

- We will assess the impact of the intervention on all official MDG targets that available data allow.
- We will analyse results separately for the following sub-groups: male and female beneficiaries; residents of Builsa and West Mamprusi districts (as of baseline administrative subdivision); and near and far control communities (as of baseline sampling stratification by distance).
- Multiple testing of MDG impacts will be corrected by false discovery rate.
- Participation in specific programme activities and targeting are analysed.
- Impacts will be estimated by the double robust method combining difference-in-differences (DD) and propensity score matching (PSM). DD and PSM will be combined using inverse probability weighting using regression analysis.
- We will estimate average treatment effects (ATE) rather than average treatment effects on the treated (ATT) in our confirmatory analysis, though the exploratory analysis will investigate impacts on specific population groups.
- Following Angrist and Pischke (2009) we will present impact estimates of: a) one of the equivalent DD fixed effect models illustrated in this document, and b) the DD lagged outcome model (also known as ANCOVA) as upper and lower bounds, respectively, of the true DD effect.
- Programme participation will be modelled differently for each observation-specific outcome using logit models.
- When data are available for more than two periods we will present estimates of average effects over the whole period as well as effects specific to each period.
- We will conduct a number of robustness checks to assess the validity of the results against the following threats to validity: seasonality; migration and attrition; differential trends in the outcomes; serial correlation; selection on the unobservables; and covariate shocks.
- We will analyse spillover effects exploiting the stratification of control villages by distance made at the sampling stage, and by using indices of social distance constructed from the social network module of the household questionnaire.

¹ Available at: www.ids.ac.uk/project/millennium-villages-in-northern-ghana-impact-evaluation

- We will analyse the impact of the intervention on a number of non-MDG outcomes that capture important domains of living standards not covered by official MDGs.
- Quantitative and qualitative teams will cooperate in formulating theories of change for specific interventions in order to assess the impact of the programme along the causal chain.
- The qualitative team will investigate in the field main anomalies found in the analysis of the data.
- We will assess changes in returns to factors produced by the intervention by structural modelling through an application of the Oaxaca decomposition method.
- We will test the presence of behaviours that are consistent with poverty trap models, in particular the impact of wage expectations on educational choices and the impact of time preferences on savings and insurance uptake.
- We will review and formulate methods for testing dynamic poverty traps that we will use once the full panel data are available.

1. Achievement of the MDGs

This analysis plan discusses our strategy for: a) a confirmatory analysis: testing whether the project goals and hypotheses are supported by the data and b) an exploratory analysis: testing the enablers and constraints that explain programme success and formulating new hypotheses about how the programme works. The confirmatory analysis is covered by the sections on: achievement of target outcomes (the MDGs); impact heterogeneity; multiple outcomes testing; programme participation; and econometric methods. The exploratory analysis is covered by the sections on: spillover effects; programme impact on non-MDGs outcomes; impact on returns to factors; and breaking the poverty trap.

The stated goal of the intervention is the achievement of the MDGs which, in the project logical framework, is instrumental to breaking the poverty trap. The first goal of the evaluation is therefore assessing to what extent the programme is improving the MDGs. The outcome indicators selected for the evaluation from the United Nations MDGs are listed and defined in Table 1. When the MDG definition adopted in the evaluation differs from the UN definition this is reported in the third column. Differences emerge mostly because of the type of data collected. Not all MDGs can be tracked with the available data. The full list of MDG indicators with the justification from exclusion from the present analysis can be found in the appendix. There are three possible reason for exclusion: 1) the indicator can only be calculated at the country level (for example, the number of seats held by women in national parliament); 2) data are not available because they were not collected in the interviews (for example, HIV prevalence); 3) data are not available because the sample size is not large enough to estimate the desired impact (for example, maternal mortality). There are a total of 28 MDG outcome indicators.

Table 1. The MDG outcome indicators

Indicator	Definition	Evaluation definition
Goal 1: Eradicate extreme poverty and hunger		
1.1 Proportion of population below \$1 (PPP) per day	The proportion of the population living in households below the international <i>poverty line</i> where the average daily consumption (or income) per person is less than \$1.25 a day measured at 2005 international prices adjusted for <i>purchasing power parity (PPP)</i>	Same definition
1.1a Proportion of population below national poverty line	The proportion of the total population living below the <i>national poverty line</i>	We use the new GLSS6 Ghanaian poverty line on a per-adult equivalent basis as calculated by GSS
1.2 Poverty gap ratio	The mean shortfall of the total population from the <i>poverty line</i> (counting the non-poor as having zero shortfall), expressed as a percentage of the <i>poverty line</i>	We use the GSS national poverty line for poverty gap computation
1.3 Share of poorest quintile in national consumption	The share of a country's national consumption or income that accrues to the <i>poorest quintile</i> (fifth) of the population	We use the 'village' total expenditure rather than the 'country' expenditure after ranking households by per-adult equivalent expenditure

1.4 Growth rate of GDP per person employed	The sum of gross value added by all resident producers in an economy, plus any product taxes and minus any subsidies not included in the value of the products	Average growth in incomes per person employed
1.5 Employment-to-population ratio	The proportion of a country's working-age population (15 and older) that is employed	The percentage of individuals older than 15 who did any work, paid or unpaid, over the previous year
1.6 Proportion of employed people living below \$1 (PPP) per day	The proportion of individuals who are <i>employed</i> , but nonetheless live in a household whose members live below the <i>poverty line</i> . Either the national poverty line or the international poverty line of \$1.25 <i>purchasing power parity (PPP)</i> per day	The percentage of the employed (any individual above the age of 15 who did any work, paid or unpaid, over the previous year) whose households is below the national poverty line
1.7 Proportion of own-account and contributing family workers in total employment	The proportion of workers in <i>self-employment</i> who do not have <i>employees</i> , and <i>unpaid family workers</i> in total <i>employment</i>	Same definition
1.8 Prevalence of underweight children under-five years of age	Children aged 0-59 months, whose weights are less than two standard deviations below the median weight for age groups in the <i>international reference population</i>	Children aged 6-59 months, whose weights are less than two standard deviations below the median weight for age groups in the WHO <i>international reference population</i>
1.9 Proportion of population below minimum level of dietary energy consumption	The proportion of the population whose food consumption is below a minimum dietary energy requirement for maintaining an acceptable minimum body size, a healthy life and carrying out light physical activity	There are two indicators: a) proportion of the population below the 'food poverty line' (or 'extreme poverty') as estimated by GSS; b) the proportion of the population whose calories consumption falls below WHO recommendations, age-sex-work specific calculated from expenditure data using national conversion factors
Goal 2: Achieve universal primary education		
2.1 Net enrolment ratio in primary education	The number of <i>children of official primary school age</i> who are enrolled in <i>primary education</i> to the total population of <i>children of official primary school age</i>	Proportion of children aged 6-11 that are attending primary school at the time of the interview
2.2 Proportion of pupils starting grade 1 who reach last grade of primary	The proportion of pupils starting grade 1 who reach last grade of primary	Probability of reaching grade 6 for a child ever enrolled in school (calculated using a survival statistical model)
2.3 Literacy rate of 15-24 year-olds, women and men	The proportion of the population aged 15-24 years who can both read and write with understanding a short simple statement on everyday life	The proportion of 15-24 year-olds who can read all the words from two simple sentences and compute two simple arithmetic operations

Goal 3: Promote gender equality and empower women		
3.1 Ratios of girls to boys in primary, secondary and tertiary education	The <i>ratio</i> of girls to boys in <i>primary, secondary or tertiary education, or Gender Parity Index</i> , is the ratio between the Gross Enrolment Ratio (GER) of girls and that of boys, for each level of education	Same definition applied to net attendance as defined in goal 2.1
3.2 Share of women in wage employment in the non-agricultural sector	Percentage of female workers in total <i>wage employment</i> in the <i>non-agricultural sector</i>	Same definition
Goal 4: Reduce child mortality		
4.1 Under-five mortality rate	The probability for a child born in a specified year to die before reaching the age of five	Calculated using the DHS synthetic cohort probability method
4.2 Infant mortality rate	The probability that a child born in a specified year will die before reaching the age of one	Calculated using the DHS synthetic cohort probability method
4.3 Proportion of 1 year-old children immunised against measles	The proportion of <i>children under one year of age</i> who have received at least one dose of <i>measles-containing vaccine</i>	Same definition
Goal 5: Improve maternal health		
5.2 Proportion of births attended by skilled health personnel	The proportion of total <i>live births</i> that are attended by a <i>skilled birth attendant</i> trained in providing life saving obstetric care	The proportion of deliveries assisted either by doctor, clinical officer, nurse, midwife, or community health worker during the last pregnancy
5.3 Contraceptive prevalence rate	The percentage of <i>women of reproductive age</i> (15-49) who are currently using, or whose sexual partner is currently using, at least one contraceptive method	Proportion of women aged 15-49 using any contraceptive method
5.5 Antenatal care coverage (at least one visit and at least four visits)	The percentage of women aged 15-49 with a <i>live birth</i> in a given time period that received <i>antenatal care</i> provided by <i>skilled health personnel</i> at least once during their pregnancy	The percentage of women aged 15-49 with a <i>live birth</i> that received at least one antenatal visit by either doctor, clinical officer, nurse, midwife or community health worker during the last pregnancy

6.3 Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS	The percentage of the population aged 15-24 that has a <i>comprehensive correct knowledge of Human immunodeficiency virus</i> (correctly identifying the two major ways of preventing the sexual transmission of HIV [using condoms and limiting sex to one faithful, uninfected partner], knowing that a healthy-looking person can transmit HIV and rejecting the two most common <i>local misconceptions</i> about HIV transmission)	Same definition based on 6 questions
6.7 Proportion of children under-5 sleeping under insecticide-treated bed nets	The proportion of children aged 0-59 months who slept under an <i>insecticide-treated mosquito net</i> the night prior to the survey	Same indicator
Goal 7: Ensure environmental sustainability		
7.8 Proportion of population using an improved drinking water source	The share of the population that uses any types of improved <i>drinking water</i> supplies	Percentage of households with access to any of the following sources of drinking water: piped into welling, yard or plot; public tap; tube well and borehole; protected dug well; protected spring; bottles; and sachet water
7.9 Proportion of population using an improved sanitation facility	The proportion of the population using an <i>improved sanitation facility</i> (a facility that hygienically separates human excreta from human, animal and insect contact)	Percentage of households with access to any of the following improved toilet facilities: flush to piped sewer system; flush to septic tank; flush to pit (latrine); ventilated improved pit latrine (VIP); and pit latrine with slab
Goal 8: Develop a global partnership for development		
8.14 Fixed-telephone subscriptions per 100 inhabitants	The sum of the active number of analogue fixed-telephone lines, voice-over-IP (VoIP) subscriptions, fixed wireless local loop (WLL) subscriptions, integrated services digital network (ISDN) voice-channel equivalents, and fixed public payphones	Percentage of people living in households reporting having a landline in the home
8.15 Mobile cellular subscriptions per 100 inhabitants	The number of <i>mobile-cellular telephone subscriptions</i> per 100 population	Percentage of people living in households reporting owning a mobile phone

Some indicators, such as literacy rates and access to safe water, are easy to calculate. Other indicators, like poverty and nutritional Z-scores, require substantial data manipulation. All these indicators can be analysed within standard t-tests of mean comparisons and ordinary least squares (OLS)/logistic regressions with project dummies. The only exceptions are the calculation of the proportion of people completing grade six, which requires a survival model, (see for example Cameron and Trivedi 2005) and the calculation of child mortality rates, which employs the method of synthetic cohort probabilities. We have designed stata codes

for the calculation of child mortality rates using the synthetic cohort probability method employed by Demographic and Health Survey (DHS) (Rutstein and Rojas 2006). The method is based on the birth history module of women interviews and can calculate mortality for different age groups and different time intervals going back at least 10-15 years from the time of the survey interview. The method adjusts for right and left censoring of observations and, via mothers' recall, exploits a large number of birth and death events. Unlike the DHS method that calculates standard errors of mortality rates using a jackknife procedure, we have developed a stata code that employs the bootstrap and that adjusts standard errors for the cluster design of the survey by re-sampling the variance (see for example Kolenikov 2010). Our method for calculating standard errors is therefore more accurate than the DHS method and produces larger standard errors than the DHS method after accounting for the two-stage cluster design of the survey.

2. Impact heterogeneity

The analysis will be conducted by disaggregating over the following sub-groups: sex of project participants; district of residence; and distance from Millennium Village (MV) areas. The reduction of gender inequalities and women empowerment is one of the MDGs, and one of the goals of the evaluation is to assess to what extent the intervention is readdressing gender imbalances and how men and women are differently affected. The project was originally implemented in the districts of Builsa and West Mamprusi (later to become three districts because of a new administrative subdivision and the creation of new districts). The original samples from Builsa and West Mamprusi are homogeneous within the district in terms of ethnic, political, and cultural characteristics and are sufficiently different from each other (different languages, social, and political organisation) to justify a separate analysis of programme impacts.² Finally, villages in the proximity of the MV areas may indirectly benefit from the intervention and an analysis of programme impact employing comparator groups at different geographic distances from the intervention is needed (see longer discussion below on the treatment of spillover effects). This subgroup classification for the analysis follows the design of the sampling frame, which stratified the sample in observations from Builsa and West Mamprusi in equal size and from 'near' and 'far' control villages in equal size, so that the disaggregated analysis by sex, district, and distance conveniently generates groups of equal size.

² On 28th June 2012, after the project and the evaluation had been designed, the Government of Ghana established 46 new districts. The study districts of Builsa and West Mamprusi were split into four new districts with the following names: Builsa North (district capital Sandema), Builsa South (Fumbisi), Mamprugu Moaduri (Yagaba), and West Mamprusi (Walewale). While CVs are equally distributed across the new four districts, the project villages are mostly in Builsa South and West Mamprusi:

	Builsa North	Builsa South	West Mamprusi	Mamprugu Moaduri
Project	1	22	7	4
Control	23	23	11	11

For theoretical and technical reasons, we will disaggregate the analysis using the old, rather than the new, district classification. First, the old district subdivision captures relevant geographic and historical institutional differences across populations. Second, the project villages are not sufficiently evenly distributed across the four districts (in particular there is only one project village in the Builsa North district) and the sample size is too small to conduct a sufficiently powered analysis across the four districts.

3. Multiple outcomes testing

Since the goal of the evaluation is to assess the impact of the intervention over several outcome indicators, the statistical analysis faces the multiple testing problem. By setting a significance level of 5% in testing hypotheses we are implicitly allowing for the occurrence of at least one rejection as the number of hypotheses tested increases. With 28 hypotheses tested, the probability of rejecting at least one null is about 75%. This is a well-known statistical problem without an obvious solution.³ One popular approach to address the multiple testing problem is the use of Bonferroni-type corrections. Bonferroni corrections change standard errors or P-values in proportion to the number of hypotheses being tested in order to make the overall test more restrictive so as to reduce the probability of finding an impact on any of the outcomes when in fact there is no impact. Bonferroni-type corrections, however, tend to be too conservative as they reduce the probability of type I errors by increasing the probability of not finding any impact when in fact there is one. The problem is compounded when the outcome variables being tested are highly correlated. When applying Bonferroni corrections it is sufficient to add more and more correlated variables, and so increasing the rigour of the test, up to the point when the null hypothesis of no impact is accepted. An additional requirement of correct application of Bonferroni adjustments is therefore that the outcomes being tested are not describing the same construct or determined by the same latent variable. A quick look at Table 1 shows that this is clearly not our case. It should also be noted that after two or more years of the project it is unlikely that the intervention does not have any impact on any of the outcomes considered as some of the MDG outcomes will be positively affected. In this circumstance the test of the null hypothesis of no impact on any variable conducted by Bonferroni adjustments does not seem appropriate.

A better option is the calculation of the False discovery rate, which estimates the fraction of null hypotheses that are wrongly rejected (Fink, McConnel, and Vollmer 2014). The false discovery rate will be our method of choice and will be applied to the multiple testing of the outcome variables in Table 1. We will also experiment with multilevel modelling of the outcomes (Gelman, Hill, and Yajima 2012). Multilevel modelling changes the point estimates in addition to the confidence intervals and provides more conservative estimates than independent hypotheses tests without reducing statistical power. The method appears particularly indicated when testing variables that are highly correlated within domains or when testing across population subgroups. However, it is much less obvious how to build a multilevel model of different impact effects on different subjects (e.g. mothers, children and farmers). Hence, we will experiment with multilevel impact estimates within limited domains. For example, this can be done for education or child health testing simultaneously for outcomes such as: Raven's matrices scores, digit span test, maths and English tests scores.

³ For a review of approaches to multiple testing offered in the statistical literature, see the review by Schochet (2008).

4. Programme participation

A fundamental step in the analysis of programme impact is the explanation of participation by villages, households, and individuals in project activities. This analysis fulfils several goals:

(1) Observing the extent of project reach by activity and its distribution across relevant groups such as: income, sex, and location. We conduct this analysis by comparing average participation rates in project-related activities by household and individuals of project and control areas (see second round report). In particular, after discussions with the project team in Bolgatanga, we look at levels of participation for the following interventions: distribution of mosquito bed nets; visit by community health workers; provision of condoms; measurement of arm circumference; advice on breastfeeding; child feeding; use of mosquito bed nets; visits to health facilities; provision of deworming tablets; provision of child supplementary food; provision of sanitary pads; membership in farmers' associations, farmers' cooperatives, farmer field schools, parent and teacher associations (PTAs), and women groups.

(2) Assessing targeting accuracy of interventions. This is performed by examining participation rates across the per-capita expenditure distribution in a comparative way between project and control areas. Differences in participation across per-capita expenditure quintiles are tested for statistical significance (see the 2nd round report).

(3) Identifying characteristics of participants (villages, households, and individuals) for the selection of comparable groups via matching methods. Characteristics of participants are obtained by running logit models (see the discussion below). We identify three levels of participation of decreasing order (see Table 2). Level 1 includes the sample of villages and households that are either in the project or the control group. Characteristics of villages and households were already analysed through the balancing tests in the baseline report. Village and household characteristics are used to run probit participation models that estimated the propensity scores employed in inverse probability weighting of observations illustrated in the next section. Participation models at this level are employed for the analysis of household-level outcomes of Table 1 such as poverty and access to improved sanitation facilities.

Level 2 includes target groups of specific interventions such as children and farmers. For example, anthropometric methods are only measured over children under-5, and enrolment rates can only be calculated over children of the relevant school age. In addition to the village and household characteristics employed at level 1, level 2 also employs individual-level characteristics such as child age and sex in the estimation of participation models.

Level 3 includes individuals or households self-selected or selected by the project within the target groups such as farmer members of cooperatives and households visiting health facilities. Exploratory analysis can be used in order to understand how the programme works. In order to do so an analysis of the impact on the treated individuals may be needed.

Table 2. Participation in programme activities and examples of characteristics determining participation

Level	Definition	Sample characteristics
Level 1	<ul style="list-style-type: none"> ▪ All sampled villages and households in project and control areas 	<ul style="list-style-type: none"> ▪ Village distance to markets, schools, and clinics ▪ Project assistance in the village (NGO, government, etc.) ▪ Village electricity ▪ Village prices (wages and food staples) ▪ Village covariate shocks and pre-baseline trends in shocks ▪ Household demographic composition ▪ Asset holdings (land, housing) ▪ Pre-baseline trends in incomes and assets ▪ Access to facilities
Level 2	<ul style="list-style-type: none"> ▪ Children aged 0 to 5 ▪ School-age children ▪ Farmers ▪ Mothers 	<ul style="list-style-type: none"> ▪ Age and sex ▪ Adult education level ▪ Relation to head of household ▪ Birth order
Level 3	<ul style="list-style-type: none"> ▪ Cooperative members ▪ Households visiting health facility ▪ Households visited by community health worker 	<ul style="list-style-type: none"> ▪ Vulnerability to shocks ▪ Time preferences attitudes ▪ Access to extension ▪ Attitudes to collective action and trust

5. Econometric Methods

5.1 Difference-in-differences

The evaluation design outlined a difference-in-differences (DD) approach for the estimation of the project impact. In a DD framework, observations in the project group are compared to observations in a control group. In the simplest setup, none of the groups is exposed to the treatment in the first period while the project group is exposed to the treatment in the second period but not the control group. Given a number of assumptions and qualifications discussed below this approach removes the biases produced by changes over time and by differences between the groups and provides an estimate of project effects.

Our sample consists of 35 project villages and 70 control villages. The project villages were selected within the two districts where the project is implemented using a one-to-one matching method based on a set of village-level variables from the 2010 census, supplemented by village-level observations collected in the field. Each project village was paired to a control village from two strata. One stratum was composed of potential controls in the vicinity of the project and the other stratum was composed of potential controls far away from the project. Hence, there are 35 project communities, 35 control communities in the vicinity of the intervention area, and 35 control communities far away from the intervention area but within the district boundaries. The oversampling of the control communities was conducted with the goal of providing an estimation of project spillover effects to neighbouring communities and of allowing the use of matching methods at the analysis stage (see Section 6.1).

In the simple standard two-period and two-group set-up, the DD effect is the difference in the change over time in the average outcomes in the project and in the control groups:

$$\delta = (\bar{y}_{P,1} - \bar{y}_{P,0}) - (\bar{y}_{C,1} - \bar{y}_{C,0})$$

where δ is the DD effect, y is the average outcome either in the project group (P) or in the control group (C) observed in the first period (0) and in the second period (1).

The same effect is conveniently measured in a regression framework by estimating the following equation:

$$y_i = a + bT_i + cP_i + dP_iT_i + e_i$$

where y is the outcome for the observation i , T is a dummy variable equal to 0 for period 1 and equal to 1 for period 2, P is a dummy variable equal to 1 if the observation is in the project group and equal to 0 if the observation is in the control group, PT is equal to 1 if the observation is both in the project group and observed in the second period. The equation estimates the following:

a is the average outcome in the control group in period 1

b is the difference in the outcomes between period 2 and period 1 in the control group (the time trend)

c is the difference between project group and control group in period 1

d is the DD effect of the project

The advantage of using regression analysis rather than simply subtracting means across groups is that regressions can include additional covariates such as, for example, initial values of household characteristics. The regression with additional covariates (X_i) is:

$$y_i = a + bT_i + cP_i + dP_iT_i + \sum_{j=0}^n g_j X_i e_i$$

There are two advantages in adding covariates. First, the covariates may balance the project and group samples, as these were not randomly obtained from an experiment. Second, the inclusion of covariates increases precision of the estimates by reducing the standard error of the coefficients. One potential problem with the use of covariates in the estimation of project effects is that most covariates are affected by the project or are themselves objectives of the intervention. Think, for example, of a DD regression of height-for-age including changes in total household expenditure. The inclusion of variables affected by the programme

will ‘absorb’ some of the project effects that would otherwise be captured by project dummies. Hence, in order to capture the programme impact with a project dummy interaction, the covariates can only include variables that are not affected by the programme and can include baseline values of the variables. This is discussed in Rosenbaum (1984), including special cases when the inclusion of post-treatment variable adjustment is feasible.

Panel data with two periods

Using panel data presents the additional advantage of removing fixed effects: unobservable variables that are time-invariant. Participation in the programme may depend on unobservable factors such as farmers’ motivation or child abilities, which are correlated with the outcomes. Similarly, project managers may target households or select communities based on unobservable characteristics such as presence of NGOs and commitment of local leaders that again are correlated with the outcomes. Panel data allow the removal of the impact of unobserved components on estimated programme effect. The estimation of DD effects with panel data can be performed in three equivalent ways that, in the two-period case, deliver the same point estimates and standard errors. We mention the three methods here for completeness, with associated stata commands, as they each may work with some data structures but not with others. In most applications we will simply employ the fixed effects model:

- The fixed effects model:

$$y_{i,t} = a_i + bT_i + dP_iT_i + e_i$$

In stata this requires setting the data as panel data with `xtset hhid T`, where `hhid` is the household id and `T` is a dummy for period 2, and then running `xtreg depvar T PT, fe`.

- The first difference model:

$$y_{i,1} - y_{i,0} = a + dP_i + e_i$$

In stata this is obtained by running the differenced variables on the project dummy using OLS: `reg d(depvar) P, cluster(clusterid)`.

- The de-meaned or **within estimator** model:

$$y_{i,1} - \bar{y}_i = b(T_i - \bar{T}_i) + d(P_iT_i - \bar{P}_i\bar{T}_i) + e_i$$

First, the means of the dependent variable, the time dummies, and the time project interaction dummies are calculated for each household. Second, the same variables are de-meaned by subtracting the household-specific means from the variable value. In stata, first calculate the mean of each variable for each household using `egen, by(group)`, then calculate the differences from the means and finally regress the de-meaned dependent variable on the de-meaned time and time project interaction dummies without the constant: `reg dm(depvar) dmT1 dmT2 dmDD, nocons`. This is equivalent to running the model using dummies for each household: in stata this can be done including slope intercepts for each household that are ‘absorbed’ by running `areg depvar T PT, absorb(hhid)`.

A different model can be employed assuming unconfoundness given the **lagged outcomes** (Imbens and Wooldridge 2009) also known as the analysis of covariance (ANCOVA) model:

$$y_{i,1} = a + by_{i,0} + dP_i + e_i$$

which in stata is simply a regression of the dependent variable in period 2 on the dependent variable in period 1 and the project dummy: `reg depvar(1) depvar(0) P`. Coefficient estimates and standard errors in this case

differ from those of the standard fixed effects DD model. Angrist and Pischke (2009) discuss the problem of correlation between the lagged value of the dependent variable and the error term in this model. Imbens and Wooldridge (2009) recommend using substantive knowledge in the choice (is the coefficient on the lag significant?) but propend for the superiority of the lagged model. Angrist and Pischke (2009) recommend using both models and interpreting the point estimates of the fixed effects model and of the lagged model as upper and lower bound, respectively, of the true causal effect. This seems sensible advice that we will follow in our work.

Panel data with multiple periods

All the methods above can be easily accommodated to the analysis of multiple periods. When analysing multiple periods, two different DD effects can be estimated. The first is the average DD effect of the programme (one project*time interaction). The second is the time-specific DD effect of the programme (several project*time interactions). For example, in the case of a three-period model the fixed effect model is (with T=0,1,2):

$$y_{i,t} = a_i + b_1T_{i,1} + b_2T_{i,2} + dP_iT_i + e_i$$

$$y_{i,t} = a_i + b_1T_{i,1} + b_2T_{i,2} + d_1P_{i,1}T_{i,1} + d_1P_{i,2}T_{i,2} + e_i$$

The second specification may be more appropriate for our study since we are interested in detecting impact that varies over time. Similar adjustments for multiple periods can be made to estimate the within-estimator model and the lagged outcomes model. On the other hand, the first difference model can create strange patterns as the number of differenced dummy variables increases.

5.2 Combining DD and matching

The simultaneous application of matching methods and DD may provide a more robust estimator. It is important to note that the project will have an impact on different socio-demographic groups at different levels: children under-5, mothers, households, farmers, etc. This means that several different participation models will be formulated for each of the levels at which the outcome is observed. For example, month of interview, age, and sex of the child will be relevant when analysing the impact on anaemia but not when analysing project impact on household poverty. Note that by balancing on the distribution of covariates when calculating DD estimates we do not mean estimating treatment-on-the-treated effects for individuals or households that are selected or self-select in specific interventions. Our goal is to combine DD and matching in the estimation of average treatment effects in order to increase the robustness of the results.

Matching can be combined with DD analysis in different ways. We list here the most common ones. Our preference is the fourth listed methodology: the use of regression weighting using inverse probability weights.

1. Use a simple combination of matching and DD (Khandker, Koolwal, and Samad 2010). First, estimate a selection equation for communities, households, or individuals in the project. Second, use the propensity score to find single or multiple matches and remove observations outside the region of common support. Finally, apply to the matched project and control observations a DD model with fixed effects (or lagged dependent variables) if using panel data, and OLS with time and treatment dummies if using cross-sectional data.
2. Combine matching and DD across all treatment groups when using cross-sections (Blundell and Costas Dias 2005), for example, when analysing effects on anthropometrics or test scores (children are leaving our panel as they grow up). They suggest the following: first, estimate a participation model in which the dependent variable is one if the observation is in the project group at follow-up. Second, match project group observations to the following three groups separately: baseline treatments, baseline controls, and

follow-up controls. Third, calculate the DD between the project and the matched project and control groups. Standard errors of the differences are calculated using bootstrap methods.

3. Combine matching and DD using semiparametric methods (Heckman, Hichimura, and Todd 1997). The methodology is rather complex and requires the following. First, run a semi-parametric regression of the outcome on the covariates in the control group and obtain the coefficients using baseline data. Partial linear regression is the recommended semiparametric method. Second, remove the predicted values of the regression from the outcome in the control group and in the project group using the same coefficients (though applied to different covariates) in the baseline data. Third, calculate the difference in the outcomes at baseline between project and control groups by calculating the differences between project outcomes calculated as above and control outcomes after matching to control observations (the recommended matching estimator is local linear matching). Remove observations outside the region of common support. Fourth, repeat the whole exercise for the follow-up data. Fifth, calculate the DD: the difference between the differences at follow-up and at baseline. Alternatively, remove the effect of covariates above from the first differences in the project group and the control group. Calculate the differences of the differences in the outcomes between the project group and the matched control group.
4. Combine matching and DD using regression weighting. Imbens and Wooldridge (2009) discuss the ‘double robustness’ of regression weighting and its superiority to either regression or matching alone. Several regression weighting methods have been suggested. Imbens suggests the following (2004). First, run a logit model of participation in the programme. Second, use the propensity scores to calculate the inverse probability weights.⁴ Third, run the standard DD regression applying the weights (this is done in stata by setting $[aweight=1/w]$ or by using weights $[sqrt(w)]$ in weighted least square – $vwls$ in stata).

A limitation of inverse probability weighting is the presence of extremely large and small weights (Stuart 2010). The estimation of participation model can result in extreme weight because some observations will have propensity scores close to one and zero. Extreme weights in turn affect parameter estimation in regression. To obviate this problem, some authors suggest trimming the extreme tails of the distribution of weights, but there is no guidance regarding the appropriate trimming thresholds. To avoid reliance on a single estimator and for robustness checking, we will also estimate effects using straight propensity score matching and then compare the results.

A key element of this empirical strategy is the formulation of a good participation model, which includes decisions on the choice of explanatory variables (determinants of participation) and the estimation method. We will employ a logit model for the estimation of participation equation (Maddala 1983):

$$y_i^* = \beta' x_i + u_i$$

where y_i^* is unobservable and what we observe is a dummy variable defined by:

$$y = 1 \text{ if } y_i^* > 0$$

⁴ The weights for the estimation of average treatment effects are:

$$w_i = \frac{T_i}{e_i} + \frac{1 - T_i}{1 - e_i}$$

where $T=1$ if the household is in the programme and $T=0$ otherwise. The term e_i is the estimated propensity score for household i . The propensity score is the conditional probability that the household is in the programme given a set of covariates. Thus, households in the programme ($T=1$) have weight $1/e$ while households in the control group are given weight $1/(1-e)$. When estimating average treatment effects on the treated, the weights are (Hirano, Imbens, and Ridder 2003):

$$w_i = T_i + (1 - T_i) \frac{e_i}{1 - e_i}$$

In other words, the weight is equal to one when the observation is in the treatment group and equal to $e/(1-e)$ when the observation is in the control group.

$$y = 0 \text{ otherwise}$$

and the likelihood function is:

$$\ln L = \sum_{i=1}^n y_i \ln \left(\frac{1}{1 + \exp(\beta^i x_i)} \right) + \sum_{i=1}^n (1 - y_i) \ln \left(\frac{\exp(\beta^i x_i)}{1 + \exp(\beta^i x_i)} \right)$$

The literature does not offer much guidance on how to choose explanatory variables in a participation model apart from the obvious recommendation that no variables affected by the intervention should be included (Caliendo and Koepeinig 2008). Heckmann et al. (1997) show that omitting important variables increases the risk of bias and that only variables that influence both the participation decision and the outcome should be included. While some semi-automated techniques for variable selection exist, the choice of the explanatory variables largely rests on sound economic theory and deep knowledge of the context in which choices analysed are made.

Models employing different sets of explanatory variables can be compared based on their ability to correctly classify the observations as participants and non-participants. The predicted classification varies with the chosen cut-off point and a commonly used method employs a cut-off of 0.5. However, this method does not work well when there are many 0 or 1 values in the sample. Heckman et al. (1997) employ a 'hit or miss' method which maximises the number of correctly identified participants by looking at estimated probability values above the sample proportion of participants. Cameron and Trivedi (2005) suggest the use of receiver operating characteristic curves to account for different levels of the cut-off points.

These methods, however, can only be applied to a subset of covariates that based on theory and context knowledge are believed to affect participation. It is very hard to list these variables in advance of starting any data analysis particularly considering that outcomes are observed at different levels: village, household, individuals and specific groups (see Table 2), each requiring a different set of explanatory variables. We can show some potential candidate variables but the final model formulation will be a result of a combination of the variable selection process by the statistical strategy outlined above and by the deepening of our knowledge of the decision process in operation in the area. A sample of likely covariates that will be included in the participation model was provided in Table 2.

A second, and not less important, issue with the choice of the model is the sensitivity of the results to different sets of covariates and assumptions. Imbens (2004) discusses a number of tests of the unconfoundness assumption and Altonji et al. (2005) propose the adoption of lower and upper bounds of treatment effects based on reasonable assumptions.

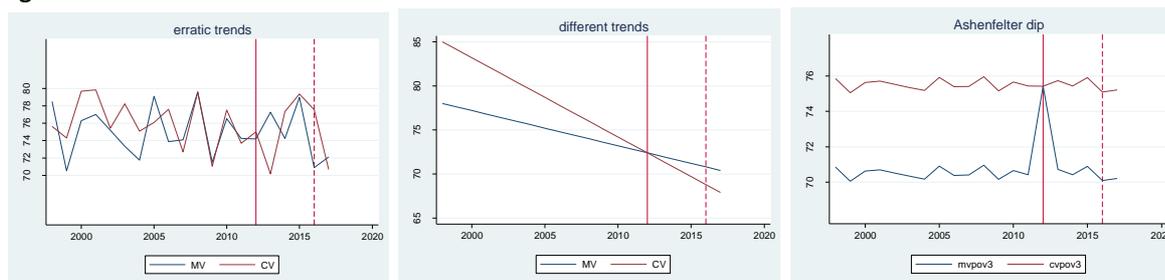
5.3 Threats to validity

Different trends

The validity of the DD approach rests on the similarity between the project and control observations in the trends of the outcome variables. The approach is valid if the changes in the outcomes observed in the control villages offer a good description of what would have happened in the project areas without the project. If the outcomes behave erratically in the absence of the programme or if there are strong and different trends in operation in the project and control areas, then DD analysis is no longer valid.

In the presence of erratic behaviour of the outcomes or of different trends in project and control areas, DD may find an impact when there is none as well as not find an impact when there is one. Moreover, the selection of the project sites may occur for specific reasons, for example, because the selected areas had been affected by a drought in the previous year, so that the following natural recovery of the target outcomes is erroneously attributed to the project. It has been suggested that MV areas are selected because of the presence of expert NGOs, the level of community organisation, and the willingness and skills of existing community leadership (Clemens and Demombynes 2011). In this case there are both observable and unobservable factors that would compromise the validity of the control group.

Figure 1. Non-MDG outcome indicators



Three examples of trends affecting DD analysis are shown in Figure 1:

- The first chart on the left is the erratic behaviour case. Suppose poverty trends in the two areas are mostly driven by covariate weather shocks like floods and droughts. The impact of these shocks on poverty is larger than any project impact and affects the two areas in very different ways. Poverty rates happen to be equal in project and control areas at the baseline (solid vertical line in year 2012) but the DD difference observed at the follow-up (dashed vertical line) would be misleading.
- The chart in the middle shows the case in which project and control areas have different trends in poverty reduction. For example, the control area might be more isolated from overall economic growth in Ghana than the project area. The lines happen to cross at the time of the baseline but the DD estimator based on the follow-up survey would again be very misleading.
- The right chart is the Ashenfelter’s dip case. Suppose the project areas are better off. Nevertheless, the government selected these particular areas because they had been affected by a severe drought the year before the intervention. Again, the DD estimator would be misleading. It would find a large programme impact while in fact the programme has had no effect and the project communities are simply recovering their normal status after the intervention.

Knowledge of these trends is essential in order for the validity of the DD approach. Trends can be used to test the validity of the approach or to improve the DD estimation. In the first case:

- Trends in the outcomes before the intervention can be inspected.
- Without information on the trends we can conduct placebo tests: testing the impact of the intervention on variables that are known not to be affected by the intervention or between areas that are both in the control group.

In the second case, trend variables collected through the questionnaires can be included among the covariates of the DD models or among the covariates of the selection models employed for matching. Table 3 shows the trend data that were collected at the baseline.

Table 3. Trend data collected by the household and community surveys in 2012

Variable	Recall period
Wage income	2010 and 2011
Microenterprise profits	2010 and 2011
Agricultural output and sales	2010 and 2011
Animal holdings	2010 and 2011
Land holdings	2010 and 2011
Prices (land, animals, fertiliser, labour, and consumer goods)	2007 and 2010
Covariate shocks	2010 and 2011
Mortality rates by cohorts (implicitly from birth histories)	Back to 1995
Education attainment by cohorts (implicitly from completion data)	Back to 1995

Selection on the unobservables

The control communities were selected from the same districts in which the project communities are located. In addition, control communities were matched to project communities based on a set of village-level characteristics. The two groups might still differ in unobservable characteristics that interact with the intervention. For example, it has been suggested that the MV area might have been chosen because of a strong presence of NGOs. The impact of the MV intervention will positively interact with the presence of NGOs. In this case the comparator group is no longer valid. The same bias could occur through several other interactions, for example, in agricultural productivity growth if agricultural extensionists were more active in the MV area and if their activity interacts with the MV project.

This type of bias can be tested and possibly reduced through matching methods if the unobservables become observable and if they are eventually observed. Cook and Campbell (1979) suggest two tests for assessing the presence of this bias: 1) if an observed variable known to interact with the unobservables varies within the project group (for example, education), then the differential impact of the project should be visible within the project group observations as well and should result in an increase in the variance of the outcomes in the project group, hence checking the size of the variances of the outcomes in the project and control groups may help detect the presence of such interacting factors; 2) plotting the baseline outcomes against the interacting variables for the project and control groups separately, a difference in the slopes between the two groups is a sign of potential bias. Meyer (1995) suggests using ‘control by systematic variation’: testing within the control group whether groups that are different in the interaction variables (for example, communities with and without NGOs) change differently over time or over other variables. Altonji et al. (2005) have developed more refined methods to assess the relative size of the impact of the unobservable which we will try to employ.

Covariate shocks

The control communities were selected from within two districts while the project communities are clusters of nearby communities. The project communities are more likely to be simultaneously affected by covariate shocks than the control communities and there are chances that covariate shocks are highly localised so as to differ between project and control communities. This is a type of clustering or village level effect. For example, many of the project communities could be affected by floods in a given year while at the same time few of the control communities are affected. In this case the validity of the control seems to be irremediably compromised. The occurrence of covariate shocks needs to be documented in order to assess the validity of the control group over time.

Shocks and serial correlation

We consider individuals i that live in J communities and are observed T times. If the i observations are correlated within communities but community-level shocks are independent across communities and time, then standard errors can be adjusted using the *cluster* option in stata for $J*T$ different clusters. Under the *cluster* option the standard errors are calculated using the Huber-White robust variance estimator, which corrects the intra-cluster correlation of the observations.

However, shocks are unlikely to be uncorrelated across time as well as they are unlikely to be uncorrelated across communities. The correlation over time has received much attention in the literature (Bertrand, Duflo, and Mullainathan 2004). If observations are serially correlated at the community level, for example, because weather shocks have long-term effects, standard errors are inconsistent. If the number of observations is sufficiently large, then a quick fix consisting of adjusting standard errors for J communities works well, but if the number of clusters is small other methods are needed (Angrist and Pischke 2009). Angrist et al. (2009) suggest in particular a) parametric correction, b) block bootstrap, and c) a modified t distribution for statistical tests. However, it is not clear what is a ‘small’ number of communities nor is it clear which method performs better among the three suggested above. Recent literature, which we will try to follow, has devised

additional methods to adjust standard errors when the number of clusters is very small and observations are not independent within the clusters (Brewer, Crossley, and Joyce 2013, Cameron and Miller 2015).

An additional problem is the correlation of observations across groups. In particular, the MV cluster is composed of 35 communities that are different but close to each other. This implies that in the project group shocks are correlated across communities. The MV cluster is not a separate entity from the surrounding area so shocks will not affect all MV communities at the same time, and they will affect control communities as well. However, the correlation across communities could be large. There might also be a correlation between project and control communities if shocks make people migrate from control villages to project villages. It is not clear how standard errors should be adjusted in these cases.

Migration and changes in the composition of the project and control groups

One problem in estimating DD regressions arises if there is a change in the composition of the groups (Angrist and Pischke 2009). For example, households may migrate from the control areas to the project areas to access project benefits. This poses a serious threat to the estimation of programme effect if, for example, malnourished children or deprived households move to the project area. Changes in outcomes observed in project and control areas can be affected by migration. Some of the migration movement can be observed as the questionnaires collect information on members moving in and out of the household. Origin and destination of migrant are known and their characteristics can be checked against average characteristics in the project and control communities. If migrants differ from average households, estimates of programme effect can be biased.

- If the analysis is conducted using cross-sectional observations the risk of bias can be large, as migration from the control group to the project group can change the distribution of the outcomes in both groups.
- If the analysis is conducted using panel observations the bias is reduced. However, it can be argued that it leaves out the changes in the outcomes among households migrating out of the control group. In addition, it could be argued that the full programme impact is not observed by panel effects, as the programme impact should include the impact on in-migrants in the project areas. In principle, impact on in-migrants could be assessed by matching in-migrants in project areas with non-migrant observations in the far away control areas (which are less likely to migrate to the project areas).

This discussion suggests that there is a need for monitoring migration flows closely in both areas in the data analysis and also with qualitative observation because in-migrants will not be interviewed by our survey unless consisting of individuals migrating into the original baseline households. Note also that this is a version of the more general 'contamination' problem that would have arisen even in experimental conditions and it is not a result of using a non-experimental approach.

Differential attrition

Households and individuals in the project and control communities may drop out of the study in different proportions and at different times. If dropouts are non-random and the characteristics of 'attritors' are correlated with the outcomes, then comparisons are biased. The presence of differential attrition can be easily checked given the wide range of covariates collected by the survey. In principle, selection models can be used to correct for attrition bias if found.

Seasonality

Baseline data were collected at different times in the project areas (April to June) and control areas (August to September). Since some of the variables vary with the seasonal cycle and data are sometimes collected using short recall periods, there is a risk of seasonality bias in the data. This risk of bias was discussed at length in Appendix H of the baseline report and we will not repeat that discussion here. Preliminary analysis

of the data and of secondary sources of similar data suggest that there is a moderate risk of bias for some health-related variables such as use of mosquito bed nets, incidence of fever and diarrhoea, and anaemia. Other outcomes, such as anthropometric measurements of children and household expenditure, are likely to be affected to a lesser extent or not affected at all. While we have indicated a number of methods via regression analysis to check the presence of a seasonal bias, there is no obvious method to estimate the size of the bias and thus correcting for the bias when estimating programme impact. In Appendix H to the baseline report we have suggested the use of an Oaxaca-type decomposition method in order to identify the ‘seasonal bias’ in the baseline differences. If the seasonal bias can be correctly identified it could then be removed from the baseline averages. It should be noted, however, that the method rests on a number of restrictive assumptions and on our ability to model the phenomenon considered using a comprehensive set of explanatory variables. In other words, the following report will flag that some of the impact on the MDGs outcomes are potentially subject to a seasonality bias. Note also that benefits of some interventions, such as the rehabilitation of roads or the installation of phone network plants, may extend beyond the villages near the project sites.

6. Exploratory analysis

6.1 Spillover effects

The programme intends to generate an imitation effect whereby other communities adopt development initiatives similar to those implemented in the MV areas. For example, a community-based health planning service clinic in a CV area may decide to adopt a system of home visits by community health workers similar to the one employed in MV areas. In addition, residents of a CV area may benefit directly from the MV services by, for example, accessing health services, school meals, markets or lower prices via road rehabilitation, and so forth. These effects can be seen in two ways. First, there are intended and unintended effects that go beyond programme targets and that need to be taken into account to provide a full estimate of programme interventions. Second, these effects compromise the validity of the control group by contamination of control observations. We will explore the occurrence of spillover effects in three ways: by using the original sampling stratification into nearby and faraway communities; by building ‘gravity models’; and by considering social distance in addition to geographic distance as a determinant of spillover effects.

Near and far away communities

The evaluation design defined two control groups based on their geographic distance from the MV area. There are two control groups of equal size: one is composed of communities surrounding the MV areas and the other is composed of communities that are far away. Provided the distribution of covariates and trends is similar across the three groups, spillover effects and the ‘true’ programme impact can be estimated in the following way:

$$y_i = a + bT_i + cP_i + dP_iT_i + gCN_i + hCN_iT_i + e_i$$

where P is, as before, an observation from the MV area, CN is an observation from nearby control areas, and observations from far away control areas are the reference group. The far away group of communities is sufficiently far to assume that residents are unaffected by the MV programme and the d coefficient measures the MV programme impact without contamination bias. The nearby communities may benefit from the programme and the coefficient h offers an estimate of the spillover effects of MVs to nearby communities.

Note, however, that the stratification by distance performed at the baseline is a general approximation based on maps and village census names that were available at the time. The survey teams are now collecting GPS locations of all villages and main facilities that will be used to check and refine the original geographic stratification. The spillover effects estimated by the equation above are a crude approximation.

Gravity models

Perhaps a better estimation of spillover effects can be obtained by making the impact of vicinity to MV areas proportional to the population size of MV villages and inversely proportional to the geographic distance. The approach is similar to the one adopted by Miguel et al. (2004) in estimating spillover effects of a deworming intervention across schools. In doing so they include in the regression model a variable for the size of the school (representing the impact of the local population density) and the number of pupils treated in nearby schools at different distances (up to 3 km and more than 3 km). However, rather than setting arbitrary distance cut-offs for the detection of spillover effects, we weight the impact of neighbouring treated population by the village distance. For example, the estimated equation could take the form:

$$y_i = a + bT_i + cP_i + dP_iT_i + gN_i + h \sum_{j=1}^n \frac{N_j}{D_{ij}} + e_i$$

where N_j is population of village j in MV areas and the summation term is the total treated population divided by the distance of each j treated village to village i , and takes therefore a different value for every village. The idea behind this specification is that spillover effects should increase with the size of the treated population but decrease with the distance of the same population.

Social distance

Geographic distance alone may not be able to explain variation in the spreading of project benefits beyond MV areas. People from neighbouring villages may have different access to project benefits (spillovers) depending on the nature and the strength of their ties with people residing in the MV villages. This could happen, for example, because they are more likely to be guests of people residing in MV areas and therefore more likely to access health or education services when needed.

The household questionnaire contains a module exploring social networks of MV and CV households. Respondents were requested to report: any important relationship with people residing in other villages (village names are reported so that the MVs can be identified); the type of relationship (at the baseline 23% were friends, 70% were relatives, and other at 7%); the strength of the relationship measured by the frequency of contacts (at baseline 50% reported having met or talked at least once a week); the hypothetical support received (at baseline 60% reported they would ask for advice, 22% for material support, and 7% for loans). This information can be combined to create an index of relevant ties with residents of MV areas, which can again be weighted by geographic distance. For example, the index could take the form:

$$\frac{kT_j}{D_{ij}}$$

where T is the number of person-ties in j MV village and k is a factor adjusting for the strength and the relevance of the tie based on the information above, and D_{ij} is again distance from village i to the MV village j as before. These indices could be included instead or in addition to the gravity model indices described above to estimate the impact of strength of social networks on the spreading of benefits outside MV areas to neighbouring areas.

6.2 Impact on non-MDG outcomes

The exploratory analysis will consider an additional set of welfare outcomes that complements those included in standard measurement of MDG achievements. Some of these indicators are as relevant, if not more relevant, than the MDGs as indicators of living standards. They include, for example, prevalence of anaemia, prevalence of stunting, cognitive ability, and scores on standard maths and language tests. A list of these outcomes by three different domains (poverty, education and health) is in Table 4, though the list is not exhaustive, particularly in the health domain, as the wealth of the data collected in the field will allow for studying other indicators. The goal of analysing these indicators is not assessing the project as a whole but investigating the project effects on the determinants of the outcomes that can help explain the observed

impact on the MDGs. In other words, the goal of this work is explaining the observed patterns in the MDGs and generating hypotheses that can be tested in other evaluations.

Table 4. Non-MDG outcome indicators

Poverty		
Average food share	Following Engel curves, food shares are an indicator of household well-being	Total household food expenditure divided by total household expenditure
Food security	Year-round adequate access to food	Three indicators: % of households not having enough food in any month over the last year; number of days households did not have enough food over the last 30 days; % of households that for any day a child went hungry the whole day during the previous year
Squared poverty gap	A measure of distributional poverty	Calculated as the poverty gap after squaring expenditure distances from the poverty line
Resilience	Population considerably affected by covariate shocks	Proportion of households considerably affected by the following shocks: drought, flood, storm, livestock death, crop losses
Agricultural income	Farmers' agricultural productivity	This includes the following indicators: output per unit of land and per unit of labour; input use (land, labour, fertiliser, and rents)
Access to credit and savings	Household's ability to borrow and save	Household access to credit over the previous year. Household savings (amount, type of accounts, animals holdings)
Household wealth	Indicator of permanent income	A composite of the value of non-productive assets, productive assets, and animal holdings
Education		
Cognitive ability	Standard tests of pattern recognition (general indicator of 'intelligence') and of short-term memory (related to attention span)	Raven matrices, forward and backward digit span tests among children aged 5 to 19 years
Quality of learning	Student's ability to read and make simple computations	There are four tests: 1) easy English: 8 questions for children from 9 to 19 who ever attended primary; 2) easy maths: 8 questions for children from 9 to 19 who ever attended primary; 3) advanced English: 20 questions for children 12 to 19 who ever attended JHS; 4) advanced maths: 25 questions for children 12 to 19 who ever attended JHS
Health		
Neo-natal, post-natal, child mortality rate	The probability for a child born in a specified year to die before reaching the age of 5. Different age ranges are related to different determinants of mortality that	Calculated using the DHS synthetic cohort probability method

	are differently affected by the intervention	
Under-nutrition indicators of stunting and wasting	Different anthropometric failures have different determinants that are differently affected by the intervention	Children aged 6-59 months, whose heights and weights are less than two standard deviations below the median weight for age groups in the WHO international reference population
Anaemia	Iron-deficiency anaemia is a major threat to child health and related to many MV interventions and outcomes	Following DHS standards, mild anaemia is calculated as the ratio of children under-5 with haemoglobin below 11 g/dL, moderate anaemia is haemoglobin below 10 g/dL, and severe anaemia is haemoglobin below 7 g/dL
Malaria	Malaria is a main determinant of mortality and of poor health among children	The child (0 to 5) is positive to either or both of two malaria tests based on parasite concentration in blood samples
Fever and diarrhoea	General indicators of morbidity strongly related to MDGs determinant and project activities	Prevalence of children under 5 reporting a fever or diarrhoea episode during the 2-weeks before the survey

6.3 Theories of change and qualitative work

The analysis described so far consisted of testing project success on a number of MDG and non-MDG outcomes. No mention was made of enablers and constraints to outcome achievements. This partly reflects the difficulty of developing impact pathways for each of a myriad of interventions and partly reflects the stress on the confirmatory role of the evaluation requested by stakeholders (testing whether MDGs are achieved or not rather than explaining why they are achieved or not achieved). However, in our exploratory work we will try to identify reasons for success or lack of success for single interventions. This in turn requires a close interaction between the work conducted by the qualitative and quantitative teams.

In particular, we are planning to work in two directions. First, we will develop micro-theories of change within the overall project by selecting those interventions that appear to be particularly relevant. For example, after field visits we have preliminarily identified the distribution of mosquito bed nets and the distribution of fertiliser as key project interventions. The qualitative and quantitative teams will develop together impact pathways for these interventions in order to investigate more closely their effectiveness along the causal chain following principles and practice of theory-based evaluation (White 2009). Second, we will conduct qualitative work to understand anomalous quantitative results with the goal of formulating hypotheses that can be tested with the available data. For example, the qualitative team can investigate the anomalous child mortality rates observed (the child mortality rate is significantly lower in MV areas, not only compared to CV areas but also compared to the rest of rural Ghana – see baseline report) and the anomalous school attendance behaviour observed (school attendance appeared to increase in MV areas but decrease in CV area – see second round report).

6.4 Returns to factors

Most of the target outcomes of the project are normally analysed in economics within household investment models. In these models households allocate resources to maximise a utility function that incorporates production, education, or health outcomes. They are, for example, wage equations (Mincer equations), profit and production functions, health (nutrition) production functions, and education models of attendance and attainment.

For example, consider a typical production function where agricultural output is a function of factors (X) such as capital, land, labour and fertiliser. Suppose this is estimated in first differences as:

$$\Delta\pi_i = \alpha + \beta(\Delta X_i) + \delta P_i + e_i$$

The problem in estimating this equation is that the programme is affecting the changes in the covariates (capital inputs and fertiliser) and possibly also changing the returns to input so that the coefficient δ does not capture the true treatment effect. An estimate of project impact in this case can be obtained by running the regression function separately for the project (1) and control (0) observations:

$$\Delta\pi_1 = \alpha_1 + \beta_1(\Delta X_1) + e_i$$

$$\Delta\pi_0 = \alpha_0 + \beta_0(\Delta X_0) + e_i$$

The changes in the covariates in the project condition can be used to predict the changes in the outcome in the control condition:

$$\Delta\pi_0^* = \alpha_0 + \beta_0(\Delta X_1)$$

And the project effect is measured by:

$$\Delta\pi_1 - \Delta\pi_0^* = (\alpha_1 - \alpha_0) + \beta_1 - \beta_0$$

which is the difference in changes in returns to factors and other unexplained differences. How we use the equations above to estimate outcomes can be explained in the following way: we test that project and control observations have the same production functions at the baseline. We can also test that the production function remains unchanged between periods in the control areas. Under these conditions, a change in factors in the project areas should generate in the project areas a movement of production/profits along the production function. If the production function shifts upwards and/or changes the slope, it must be the result of changes in returns and in other unexplained project factors. The shift in the production function is what is measured by $\Delta\pi_1 - \Delta\pi_0^*$, which is the difference between the observed change in the project areas and the predicted change using the parameters in the control areas. All this holds true under the assumption that project allocation is as good as random after controlling for the observables.

This analysis can be taken a step further into a full Oaxaca decomposition by building counterfactuals in the project and control areas:

$$\Delta\pi_0^1 = \alpha_0 + \beta_1(\Delta X_0)$$

$$\Delta\pi_1^0 = \alpha_1 + \beta_0(\Delta X_1)$$

The project effect is then:

$$\begin{aligned} \Delta\pi_1 - \Delta\pi_0 &= \Delta\pi_1 - \Delta\pi_0^1 + \Delta\pi_0^1 - \Delta\pi_0 \\ &= (\alpha_1 - \alpha_0) + \beta_1(\Delta X_1 - \Delta X_0) + (\beta_1 - \beta_0)\Delta X_0 \end{aligned}$$

which decomposes into a change in factors, a change in returns to factors, and a change to an otherwise unexplained component. The advantage of this formulation is that it allows, for example, the estimation of changes in returns to factors such as fertiliser in a production function, education in a health-nutrition function, and per capita expenditure in a school attainment function. When applied over several time periods it shows the changes in returns to factors over time that create poverty traps and how the poverty trap can be broken, which is discussed below.

6.5 Breaking the poverty trap

The ultimate goal of the MV project is to break the poverty trap for which the achievement of MDGs is instrumental. The best way to start an investigation of poverty traps is to look at the distribution of incomes

and wealth. One view of poverty traps is based on the idea that there are multiple equilibria, whose presence should lead to a clustering of households or communities around ‘poor’ and ‘rich’ areas (bimodal distribution). This is unlikely to occur in our study area where most people are poor so that traps are unlikely to be visible on a cross-sectional basis. The poverty trap is more likely to be a ‘low-equilibrium’ trap, which is common to the project and the control groups alike. The evaluation might be able to observe escapes from poverty traps by the entire MV area, by some MV communities, or by some households or individuals within MV communities. This should result in patterns of growth in income or assets that are divergent in MV areas with respect to the control areas. There are at least three ways in which these patterns can be investigated:

1. The DD analysis of impact on incomes, expenditure, profits, and asset accumulation will show whether the impact increases over time and for which households in particular. The structural modelling analysis in the previous section will show whether returns to factors are increasing in the project areas over time.
2. A second approach consists of measuring consumption dynamics directly by exploiting the panel structure of the data. There is ample literature which has tried to estimate dynamic poverty traps using short panels and household-level data. The analysis consists of regression models of poverty (or other variables) on lagged values using high order polynomials (Jalan and Ravallion 2002), non-parametric methods (Lybbert et al. 2004), or semi-parametric methods (Antman and McKenzie 2007) to detect non-linearities. There are also some applications of this approach to project evaluations in Jalan and Ravallion (1998), Ravallion and Chen (2005), and Chen et al. (2009). Econometric approaches to accomplish the estimation of non-linearities are quite sophisticated and will become possible once at least three waves of data are collected. We will explore these methods in the future.
3. Barret and Carter (2013), reflecting on the theoretical and technical limitations of models of income dynamics mentioned above, have recommended the adoption of a less ambitious approach to detect poverty traps. Rather than trying to observe non-linearities, researchers can try to observe behaviours that are consistent with theoretical models of poverty traps. Poverty traps can originate from ‘irrational’ behaviours, lack of information, or missing markets. This approach can be adopted even in the analysis of cross-sections and is one that we will follow in our study. In particular we will consider the following determinants of poverty traps:

The impact of wage expectations on educational choice. School attendance and years of schooling are usually explained as parents’ decisions based on a comparison of the benefits and costs of schooling. This cost-benefit analysis can be presented as the solution of a household maximisation model (Glewwe 1999). Benefits consist of higher productivity and wages while costs consist of the opportunity costs of schooling (time spent working in the home or in the farm), school costs (such as school fees, uniforms, PTA contributions, and stationery), and interest rates. Markets remunerate skills rather than schooling and therefore factors affecting learning also affect parents’ schooling decisions, in particular, child ability and motivation, school quality and parents’ background. Ability to borrow or save and fund schooling depends on initial wealth and current income. Demographic factors, such as number of children, spacing between births, sex of the child and birth order, also matter because household resources have to be allocated among different children. Note that while some of the above factors are observed by parents, others are not. In particular, parents may have wrong expectations about wages earned by educated children and uncertainty can decrease the average expectations. Finally, discounting of future benefits (impatience) and parents’ time horizons also have an impact on the definition of future benefits. Our survey is unique in that it includes attempts to measure parents’ time discounting, parents’ survival probabilities, parents’ and children’s wage expectations for different levels of schooling, and children’s innate abilities. This set of variables will allow the estimation of attendance models (probit) and school attainment models (ordered probit) as well as assess the impact on schooling decisions of impatience and wage expectations and to what extent the project affects schooling through changes in these factors.

The impact of ‘impatience’ on saving decisions including medical insurance. Since 2003, Ghanaian households have had access to a wide range of health services after registration with the National Health Insurance Scheme (NHIS). The scheme was shown to have a positive impact on access to services and

mothers' health (Mensah, Opong, and Schmidt 2010). Nevertheless, enrolment in the scheme remains low (less than 50% of the population in Northern Ghana). Acharya et al. (2013) reviewed the determinants of enrolment identified in the literature. They conclude that, despite enrolment being the single most important precondition for project success, reasons for non-enrolment are not well understood. Few models of enrolment have been estimated and they are empirical relationships not based on a theoretical model of household behaviour. We propose to formulate a theoretical model of household enrolment in the insurance scheme and to test the relevance of variables such as rates of time preference and survival expectations using a simple discrete choice model. Note that enrolment in NHIS is promoted by other interventions such as the Livelihood Empowerment Against Poverty programme in the area and this will be taken into account in the analysis.

References

- Acharya, A., S. Vellakkal, F. Taylor, E. Masset, A. Satija, M. Burke, and E. Shah. 2013. "The Impact of Health Insurance Schemes for the Informal Sector in Low and Middle Income Countries: A Systematic Review." *World Bank Research Observer* 28 (2):236-266.
- Altonji, J. G., T. E. Elder, and C. R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1):151-184.
- Angrist, J. D., and J-S. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.
- Antman, F., and D. McKenzie. 2007. "Poverty Traps and Nonlinear Dynamics with Measurement Error and Individual Heterogeneity." *Journal of Development Studies* 43 (6):1057-1083.
- Barrett, C. B., and M. R. Carter. 2013. "The Economics of Poverty Traps and Persistent Poverty: Empirical and Policy Implications." *Journal of Development Studies* 49 (7):976-990.
- Bertrand, M., E. Duflo, and S. Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119 (1):249-275.
- Blundell, R., and M. Costas Dias. 2005. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources* 44 (3):565-640.
- Brewer, M., T. F. Crossley, and R. Joyce. 2013. "Inference with Difference-in Differences Revisited." *IZA Discussion Paper* 7742.
- Caliendo, M., and S. Koepeinig. 2008. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22 (1):31-72.
- Cameron, A. C., and D. L. Miller. 2015. "A Practitioner's Guide to Cluster Robust Inference." *Journal of Human Resources*.
- Cameron, A. C., and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- Chen, S. H., R. Mu, and M. Ravallion. 2009. "Are there lasting impacts of aid to poor areas?" *Journal of Public Economics* 93 (3-4):512-528.
- Clemens, M.A., and G. Demombynes. 2011. "When does rigorous impact evaluation make a difference? The case of the Millennium Villages." *Journal of Development Effectiveness* 3 (3):305-339.
- Cook, D.C, and D.T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally Publishing Company.
- Fink, G., M. McConnel, and S. Vollmer. 2014. "Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures." *Journal of Development Effectiveness* 6 (1):44-57.
- Gelman, A., J. Hill, and M. Yajima. 2012. "Why We (Usually) Don't Have to Worry About Multiple Comparisons." *Journal of Research in Educational Effectiveness* 5:189-211.
- Glewwe, P. 1999. *The Economics of School Quality Investments in Developing Countries: An Empirical Study of Ghana*. London Macmillan.
- Heckman, J. J., H. Hichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 64 (4):605-654.
- Hirano, K., G. W. Imbens, and G. Ridder. 2003. "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica* 71:1161-1179.
- Imbens, G. W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* 86 (1):4-29.
- Imbens, G. W., and J. M. Wooldridge. 2009. "Recent Developments in the Econometrics of Impact Evaluation." *Journal of Economic Literature* 47 (1):5-86.
- Jalan, J., and M. Ravallion. 1998. "Are there dynamic gains from a poor-area development program?" *Journal of Public Economics* 67 (1):65-85.
- Jalan, J., and M. Ravallion. 2002. "Household Income Dynamics in Rural China." *Wider Discussion Paper* No. 2002/10.
- Khandker, S. R., G. B. Koolwal, and H. A. Samad. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington DC: The World Bank.
- Kolenikov, S. 2010. "Resampling variance estimation for complex survey data." *Sata Journal* 10 (2):165-199.

- Lybbert, T. J., C. B. Barrett, S. Desta, and D. L. Coppock. 2004. "Stochastic wealth dynamics and risk management among a poor population." *Economic Journal* 114 (498):750-777.
- Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Mensah, J., J. R. Oppong, and C. M. Schmidt. 2010. "Ghana's National Health Insurance Scheme in the Context of the Health MDGs: An Empirical Evaluation Using Propensity Score Matching." *Health Economics* 19:95-106.
- Meyer, B. D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics* 13 (2):151-161.
- Miguel, E., and M. Kremer. 2004. "Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1):159-217.
- Ravallion, M., and S. H. Chen. 2005. "Hidden impact? Household saving in response to a poor-area development project." *Journal of Public Economics* 89 (11-12):2183-2204.
- Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment." *Journal of the Royal Statistical Society. Series A (General)* 147 (5):656-666.
- Rutstein, S.O., and G. Rojas. 2006. *Guide to DHS statistics: Demographic and Health Survey Methodology*. Calverton, MD: ORC Macro.
- Schochet, Peter Z. 2008. "Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations." *Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education*.
- Stuart, Elisabeth A. 2010. "Matching Methods for Causal Inference." *Statistical Science* 25 (1):1-21.
- White, Howard. 2009. "Theory-based impact evaluation: principles and practice." *Journal of Development Effectiveness* 1 (3):271-284.