UK Department for International Development

THE IMPACT EVALUATION OF THE MILLENNIUM VILLAGES PROJECT:

ANNEX A: PRELIMINARY STATISTICAL ANALYSIS

Authors: Edoardo Masset

January 2018

Results in development

Disclaimer

The views expressed in this report are those of the evaluators. They do not represent those of the UK Department for International Development or of any of the individuals and organisations referred to in the report.

UKaid
from the British people

Report

Northern Ghana Millennium Village impact evaluation: endline impact analysis

Table of contents

List of figures

List of tables

Acronyms

AEA	Agricultural Extension Agent
ANCOVA	Analysis of Covariance
CEW	Community Education Worker
CHW	Community Health Worker
CV	Comparison Village
DD	Difference-in-Difference
DHS	Demographic and Health Survey
GES	Ghana Education Service
GSS	Ghana Statistical Service
GLSS	Ghana Living Standards Survey
IPW	Inverse Probability Weighting
ISSER	Institute of Statistical, Social and Economic Research
MDG	Millennium Development Goal
MPI	Multidimensional Poverty Index
MV	Millennium Village
MVP	Millennium Villages Project
NGO	Non-Governmental Organisation
OECD	Organisation for Economic Co-operation and Development
WHO	World Health Organization
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
PTA	Parent-Teacher Association
SMC	School Management Committee
SDG	Sustainable Development Goal
UNDP	United Nations Development Programme

1. Introduction

This endline report is the final assessment of the impact of MVP. The report builds on previous reports. Some of the conclusions and the data from the previous reports are presented again; others have been reanalysed and modified in light of the information emerging from the last round of data. The statistical analysis follows closely the pre-analysis plan both in the methods used and in the topics covered.

We start in Section 2 by summarising the evaluation design and the stratification of the control group by distance to the MV villages. Newly obtained GPS locations allowed a better definition of the strata. We then present the statistical models used to assess impact and discuss some technical issues that were not resolved in previous reports. In Section 3, we present the datasets and an overall assessment of the quality of the data and their suitability for a difference-in-difference (DD) analysis. We cover issues such as attrition, differences in trends and changes in household composition and household size, showing that they do not compromise the validity of the analysis. In Section 4, we present take-up rates for most activities promoted by MVP and analyse the characteristics of households and individuals taking part in project activities and beneficiaries. Section 5 assesses the impact of MVP on its final outcomes, the Millennium Development Goals (MDGs), using a 'dashboard approach'. In Section 6, we assess the impact of MVP on an aggregate measure of poverty using the multidimensional Oxford deprivation index. In Section 7, we discuss the theory of change (TOC) of MVP. MVP was not designed with a proper TOC, and the project activities changed over the course of the project by means of a learning-by-doing process that was built into the project design. We illustrate the difficulty of explaining the how and why of project impact in these circumstances and a justification for the evaluation strategy adopted. In Section 8, we present the impact of MVP on expenditure, income and savings and try to reconcile the impact observed on these aggregates within a unified conceptual framework. In Sections 9-11, we present detailed impacts by sector (in agriculture, health and education) on non-MDG outcomes and other intermediate indicators, in order to uncover causal mechanisms of impact. In Section 12, we discuss impacts not covered in other sections, such as migration, water access and vulnerability. Finally, Section 13 offers a quick summary of the main conclusions.

2. Evaluation design

2.1. Summary of evaluation design

The evaluation uses a mixed methods approach to impact evaluation (Masset, Acharya, Barnett, & Dogbe, 2013). At the core of the quantitative methodology is a DD approach that compares changes in outcomes in the MV areas to changes in the same outcomes in a comparison group (the 'control villages' (CV areas). Provided some conditions are met, DD isolates the MVP impact on outcomes from effects of other variables changing over time.

In preparing the design of this impact evaluation, a number of alternatives were considered. A randomised trial was considered impractical because the intervention is implemented in a cluster of geographical adjacent villages. The possible randomisation design in this context would have been the random selection of a 'control cluster'. A randomised control cluster, however, would be at risk of being different from the project cluster and being subject to covariate shocks distinct from those affecting the project cluster. The evaluation team therefore focused its effort around the best way to build a quasi-experimental control group. In the end, the selection of control villages by means of matching aggregate village characteristics, within the district in which MVP was implemented, and further matching of project and control households at the analysis stage (on household characteristics) within a DD approach, was considered the next best feasible approach. For a more detailed explanation, see Masset et al. (Masset et al., 2013).

Our sample consists of 35 project villages and 68 control villages. All MV villages were included in the study regardless of size. The average size of a village in the project area is 111 households (758 individuals), ranging from a minimum of 8 households (40 individuals) to a maximum of 527 (corresponding to 3,761 people). The total population affected by the intervention and residing in the 35 villages at the baseline is 3,901 households (corresponding to 26,591 individuals). The control 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 characteristics obtained from the 2010 population census, supplemented by village-level observations collected in the field.[1] 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 from the project. Hence, there are 35 project communities, 34 control communities in the vicinity of the intervention area and 34 control communities far away from the intervention area but within the district boundaries. The oversampling of the control communities (two control villages per each project village) was conducted with the goal of providing an estimation of project spill-over effects to neighbouring communities and with the secondary goal of building a large control sample that would allow further use of matching methods at the household level at the analysis stage (see Section 6.1, Analysis Plan).

The size of the sample was identified by performing power calculations to identify a range of project effects on a number of outcome variables (Masset et al., 2013). Given the large diversity in population size of the project villages, with the number of households running from 8 to 527, it was decided to draw the sample proportionally to village population size rather than drawing a fixed number of sample households from each village. The latter approach is the norm in impact evaluations and produces results that generalise to a population of villages. In our case, we are estimating the impact of the intervention on individuals rather than on villages, and the results generalise to a population of individuals rather than to a population of villages. If required, village-level results could be recovered by re-weighting the observations appropriately, but we believe that a population-level impact is preferable to a village-level impact. The two may differ if impacts are correlated to village size. For example, if the project is more effective in small villages, for example because it allocates a fixed financial disbursement to each village, the usual practice of drawing fixed samples from each village overstates the impact of the intervention on the population. A sample drawn proportionally to the size of the villages, on the other hand, better describes population impacts when these vary across village size (see Raudenbush & Bloom, 2015 for a discussion).[2]

2.2. Stratification

The selection of control villages was conducted within strata in the Builsa and West Mamprusi districts. The strata were based on the distance of the control villages from the MVP locations. Two strata of control villages were defined: the control villages near MVP and the control villages far from MVP. The distance-based stratification of the comparison sample of villages was devised to assess spill-over effects. The idea was that comparison villages nearby the MV villages would be somewhat exposed to the intervention while faraway villages would not benefit from the intervention. Given this set-up, a difference in outcomes between 'near' and 'far' villages could be interpreted as suggestive of spill-over effects emanating from MVP locations.

At the time of the selection of the comparison villages, no distance data between locations were available. Lacking data on distances between villages, or maps from which distances could be calculated, we decided to sample 'near' villages from area councils (sub-district administrative subdivisions comprising several village) in which the programme was implemented and contiguous area councils. The remaining area councils (non-contiguous to the councils in which the project was being implemented) were used to sample 'far' control villages. We did this separately for Builsa and West Mamprusi. The Builsa district was split in two areas: 1) Chansa, Fumbisi, Kadema, Kanjarga and the Ysobsa electoral area in Wiaga and 2) Chuchuliga, Sandema, Siniensi and Wiaga (with the exception of Ysobsa). Most localities in this area are far from the MV sites. In West Mamprusi we selected 'near' villages from the following areas: Gbmisi/Wulugu, Kpasenkpe, Kunkwa, Wungu, Yagaba and Yzesi and 'far' control villages from Kparigu, Gbmisi, Walewale and Kubore.

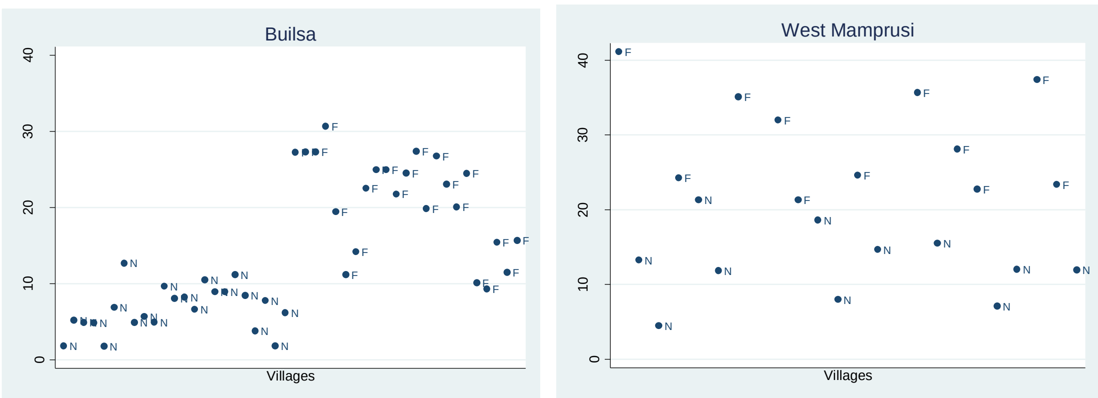
[1] The project districts are Builsa and West Mamprusi with the geographical boundaries defined in 2012. The Builsa district was later sub-divided by the government into two administrative districts so that the project is currently implemented in three districts rather than two. Subdivision of the West Mamprusi district did not affect the project as all project villages were located in one of the new districts in the western area of the original West Mamprusi.

[2] In Section 5 on the impact of MVP on the MDGs, we briefly discuss a comparison of the impacts of MVP at the individual level and at the village level. The results and a full discussion are reported in Appendix B.

Some ad hoc changes had to be made in the field as some of the selected locations could not be found with the names reported in the census while others were found outside the expected area councils. In these cases, the villages had to be replaced by other closest-matched comparison villages. During the baseline GPS coordinates were collected and we became able to calculate actual distances between villages.[3] The calculation of distances between villages and field visits highlighted a few obvious errors. The villages of Zukpeni and Tantala in West Mamprusi were not 'near' MV villages, while the villages of Zangu-vuga and Nayoku (always in West Mamprusi) were clearly not 'far' from the MV villages of Nabari and Silinga. The villages were therefore swapped.

The charts in Figure 1 plot the distance of each comparison village to the nearest project village for the Builsa and West Mamprusi districts separately. The comparison villages assigned to the faraway strata are indicated by a capital F, while the comparison villages assigned to the near strata are indicated by capital N. Distances in kilometres are reported on the vertical axis.

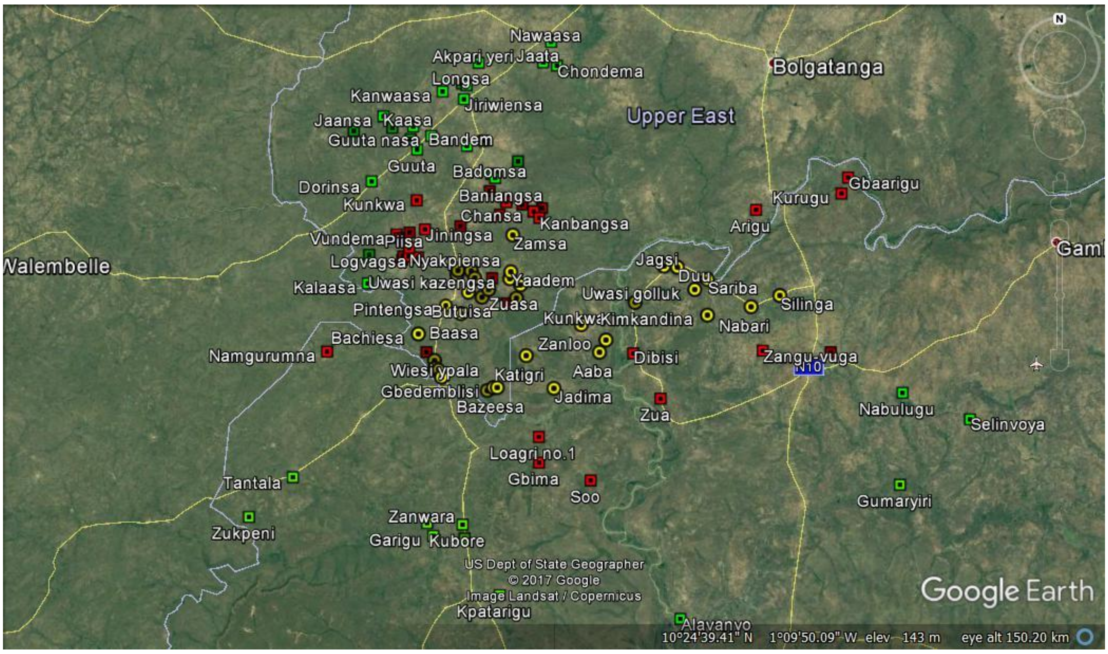
Figure 1 Distances between MV and CV villages in Builsa and West Mamprusi



Communities in West Mamprusi are geographically dispersed and relatively few communities are within 10km distance from project villages. A distance cut-off of 20km splits the sample of comparison villages in two equal parts. Villages are more contiguous in the Builsa district and there are several villages under the distance of 10km from MV villages. It also appears there are several communities located between 10km and 20km from MV villages. These villages were labelled as 'far' though, to be consistent with the classification made in West Mamprusi, they should be categorised as 'near'. However, these eight communities are apparently near the MV cluster of villages only because they are close to one single project village in Northern Builsa (Zamsa). If this single northern community were removed from the MV cluster the distance of the control community from the cluster would be much larger. The stratification of the comparison villages is illustrated in the map in Figure 2. The map reports the project villages in yellow. The red dots are 'near' comparison villages and the green ones are 'far' comparison villages.

[3] GPS longitude and latitude coordinates were collected by survey supervisors in what appeared to be the 'centre' or focal point of the community during the interviews. We converted longitudes and latitudes in northings (N) and eastings (E) and we calculated a full matrix of distances between all villages using the following formula: distance in Km=sqrt((N1-N2)+(E1-E2))/1000.

Figure 2 Map of MV and CV villages



Note: Yellow markers are MV villages, red markers are 'near' control villages and green markers are 'far' control villages.

There is always some degree of arbitrariness in setting a distance cut-off point beyond which the project is considered able or unable to produce any effect. Other evaluation designs have adopted distance cut-offs much shorter than 20km (Benjamin-Chung et al., 2015). However, it should be noted that communities are very geographically dispersed in the study area, particularly in West Mamprusi, where villages neighbouring at a distance of less than 5km do not exist. On the other hand, there have been reports from field visits of people travelling considerable distances from comparison areas to access the services offered by MVP.

2.3. Methodology

Project impact is estimated using a difference-in-difference (DD) analysis: the difference in the change over time in the average outcomes between the project and in the comparison groups. In the simple standard two-period and two-group set-up, the DD effect is:

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

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

We calculate DD effect using regression analysis. We use different regression models depending on whether panel data are available on households or individual observations. When panel data are available we employ **fixed-effect** and **lagged dependent variable model** (or ANCOVA – analysis of covariance). In a number of instances panel data are not available, like when estimating impact on nutrition, child mortality or education tests. For example, children who were under five at the baseline are no longer under five after five years and cannot be tracked to analyse nutrition or mortality. Similarly, many, if not all, children who were tested in English and maths at the baseline are no longer eligible for testing after five years. In all these cases we employ a **cross-sectional model**.

The **cross-sectional regression (t=0,1)** is also the simplest:

$$y_i = a + bT_i + cP_i + dP_iT_i + \sum_{j=1}^{n} g_j X_{ji0} + e_i$$

where y is the outcome observed 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. The interaction of the project variable and the time variable (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; finally d is the required DD effect of the project. The X_i are ($j=1,\ldots,n$) covariates that improve the balance between the project and the control group samples. The samples of project and control observations were not randomly obtained from an experiment, and including the covariates in the regression model increases the precision of the estimates by reducing the standard error of the coefficients (g_i).[4]

When panel data are available we use a **fixed effects model** to remove the impact of fixed effects: time-invariant unobservable determinants of the outcomes such as, for example, farmers' motivation or children's abilities. The fixed effect model is:

$$y_{it} = a_i + bT_{it} + dP_iT_i + \sum_{j=1}^{n} g_j X_{itj} + e_{it}$$

The covariates in this case are time-varying variables that are not affected by the project such as the occurrence of drought or other shocks and household composition.

With panel data we also employ the **lagged dependent variable model** (Imbens & Wooldridge, 2009), also known as the analysis of covariance (ANCOVA) model:

$$y_{i1} = a + by_{i0} + dP_i + \sum_{j=1}^{n} g_j X_{ji0} + e_i$$

which is simply a regression of the dependent variable in period 2 on the dependent variable in period 1 and a project dummy in addition to the usual baseline covariates.

These models can be expanded to include multiple time periods and for completeness we report below the model specifications employing five time periods. For each of the three models above we report the specification estimating the average project effect over the five-year period and the specifications estimating four year-specific project effects. The latter models estimate the impact of the intervention in each year with respect to the baseline while the former estimates the average impact of the intervention across four years with respect to the baseline.

The five-period **cross-sectional models** (t=0,1,2,3,4) are:

$$y_i = a + \textstyle\sum_{t=1}^{4} b_t T_{it} + cP_i + dP_iT_i + \sum_{j=1}^{n} g_j X_{ji0} + e_i$$

$$y_i = a + \textstyle\sum_{t=1}^{4} b_t T_{it} + cP_i + \sum_{t=1}^{4} d_t P_i T_{it} + \sum_{j=1}^{n} g_j X_{ji0} + e_i$$

The five-period **fixed effects** models (t=0,1,2,3,4) are:

$$y_{it} = a_i + \sum\nolimits_{t=1}^{4} b_t T_{it} + dP_iT_{it} + \sum_{j=1}^{n} g_j X_{jit} + e_{it}$$

$$y_{it} = a_i + \sum\nolimits_{t=1}^{4} b_t T_{it} + \sum\nolimits_{t=1}^{4} d_t P_i T_{it} + \sum_{j=1}^{n} g_j X_{jit} + e_{it}$$

The five-period **lagged** models (t=1,2,3,4) are:

[4] 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 baseline characteristics or variables that are not affected by the programme (Rosenbaum, 1984).

$$y_{it} = a + by_{i0} + \sum\nolimits_{t=2}^{4} c_t T_{it} + dP_i T_{it} + \sum_{j=1}^{n} g_j X_{ji0} + e_i$$

$$y_{it} = a + by_{i0} + \sum\nolimits_{t=2}^{4} c_t T_{it} + \sum\nolimits_{t=1}^{4} d_t T_{it} + \sum_{j=1}^{n} g_j X_{ji0} + e_i$$

The comparator villages were identified by matching villages to project villages using a propensity score. The propensity score was estimated using village-level characteristics obtained from census data and from field visits (Masset et al., 2013). In order to remove any remaining observable differences in characteristics between the project and the control group we further employ matching methods at the household and individual level in the estimation of the project effects. In doing so we follow matching on sub-classification of the propensity score as recommended and outlined by Imbens and Rubin (2015). The sub-classification procedure builds groups of project and control observations with a similar propensity scores and then employs the regression models outlined above to estimate project effects within groups. The group-level effects are then averaged proportionally to the group size to obtain the overall impact across groups. In some cases, because of modelling complexity (for example in the estimation of infant mortality), the sub-classification method cannot be employed and we use inverse probability weighting (IPW). In this method, all observations in the regression models above are weighted by the propensity score. All charts presented in the paper also weight observations by the propensity score using IPW. The details, justification and sensitivity analysis of the matching methods adopted are illustrated in Appendix A.

All regressions are estimated using OLS, including cases in which the depended variable is binary such as, for example, students' enrolment or poverty. The linear probability model (OLS) is not necessarily the best model for the estimation of binary outcomes. In particular, linear probability models have three main problems (Collet, 2002; Greene, 2011; Maddala, 1983)1) the variance is heteroskedastic, 2) they assume normality when the distribution is binomial and 3) they predict the outcome outside the permissible probability interval (0,1). In what follows we briefly discuss the main issues in employing OLS with binary outcomes; we explain why the issues have limited relevance in our case and finally we justify the use of OLS in our empirical application.

First, linear probability models are heteroskedastic. It can be easily shown (see for example Maddala, 1983) that the variance of the dependent variable is not constant and depends on the value of the proportion of successes. The variance is larger for probability values around the middle of the distribution (0.5). In this context, OLS estimation is unbiased and consistent but not efficient because there are estimators with lower variance. Heteroskedasticity is a minor complication that can be addressed by iterative estimation procedures. In our application heteroskedasticity is even less of an issue because we are estimating standard errors by bootstrapping over a very large number of sample replications.

Second, OLS (and related statistical tests) assume normality of the distribution of the error term while the dependent variable of a binary model follows a binomial distribution. When the sample is small and the values of the probability are close to 0 or 1, the distribution is skewed to the left or to the right and assumptions about probability and testing break down. However, by the central limit theorem the distribution of a binomial variable becomes normal in large samples. In our application the binomial distribution of the dependent variable is not an issue because our sample is sufficiently large for the dependent variable to follow the normal distribution. Finally, the biggest problem with linear probability models is that they make predictions that fall outside the permissible probability interval (0,1). In addition, if a large share of the dependent variable is close to 0 or 1, the regression coefficient is biased (see Pindyck & Rubinfeld, 1998). Non-linear models, such as logit or probit, constrain the probability to lie in the interval (0,1), in addition to addressing the heteroskedasticity and normality problems outlined above. Hence there seem to be plenty of benefits in using non-linear models.

Not all is good with non-linear models, however. In particular, the coefficients of a logit or probit model calculates 'odds ratios', which are quantities difficult to interpret. When the probability of success is very small, the odds ratio is very similar to the risk ratio, which is very easy to understand, but this is not the norm. For the results of the model to be better understood, it is preferable to calculate 'marginal effects'. Marginal effects are the derivatives of the logit function with respect to a particular variable. Because the model is non-linear, the derivative varies across the observations and the practice is calculating the average of the derivative: the 'average marginal effect'. When the model includes only explanatory dummy variables the regression coefficients of the linear probability model and the average marginal effects of the non-linear model are identical. But when there are interaction terms, as for example when estimating DD, the marginal effects cannot be calculated from the derivative of the interaction term (Ai & Norton, 2003). The correct marginal effect has to be calculated from the cross derivatives with respect to the interacted variables. The calculation of cross derivatives from logit models to derive correct DD effects is not overly complicated but requires time. In our matching algorithm we would be estimating logit models for 8 sub-samples of the propensity score and perform at least 500 sample bootstrap replications to calculate the standard errors, which would take an interminable time.

We decided to use OLS for the estimation of project effects for the following reasons. First, the DD effect is simply the coefficient of an interaction term between project and time variables and cross derivatives do not need to be estimated as in the case of non-linear models. Second, the coefficient is easily interpreted as a percent impact difference and the difficulties in interpreting changes in odds ratios are avoided. Finally, we conducted some tests comparing estimations by logit and OLS on a number of outcomes and found the differences in the estimated coefficients and standard errors to be negligible.

3. Data

This section describes the characteristics of the data collected, its quality and its suitability for assessing project impact. The validity of a DD approach rests on the assumption that project and comparison groups are similar on the levels as well as on the trends of the characteristics determining the outcomes. Attrition, different trends and changes in household composition can all affect the validity of DD estimation. Finally, as in all empirical research, the validity of the conclusions largely depends on the quality of the data collected in the field.

3.1. Data collection

The baseline survey aimed at collecting data from a sample of 755 households in the MV villages and 1,496 households in the CV villages. These sample sizes were identified to detect impacts of an acceptable size for key outcomes through power calculations (Masset et al., 2013). The size of the comparison group was set to be twice the size of the project group for two reasons: 1) to stratify the impact of the intervention by distance thus identifying spill-over effects and 2) to be able to perform matching of observations at the household level in order to further improve the comparability of the two samples. In every survey round, the same households were selected for the interview, though at each round not all selected households were found. As a result, the samples vary at each survey round while the sample of panel households decreases over time. We decided to follow this approach, rather than only following panel households over time, because a number of impacts, such as mortality, nutrition and education tests, are estimated over cross-sections and therefore benefit from larger samples.

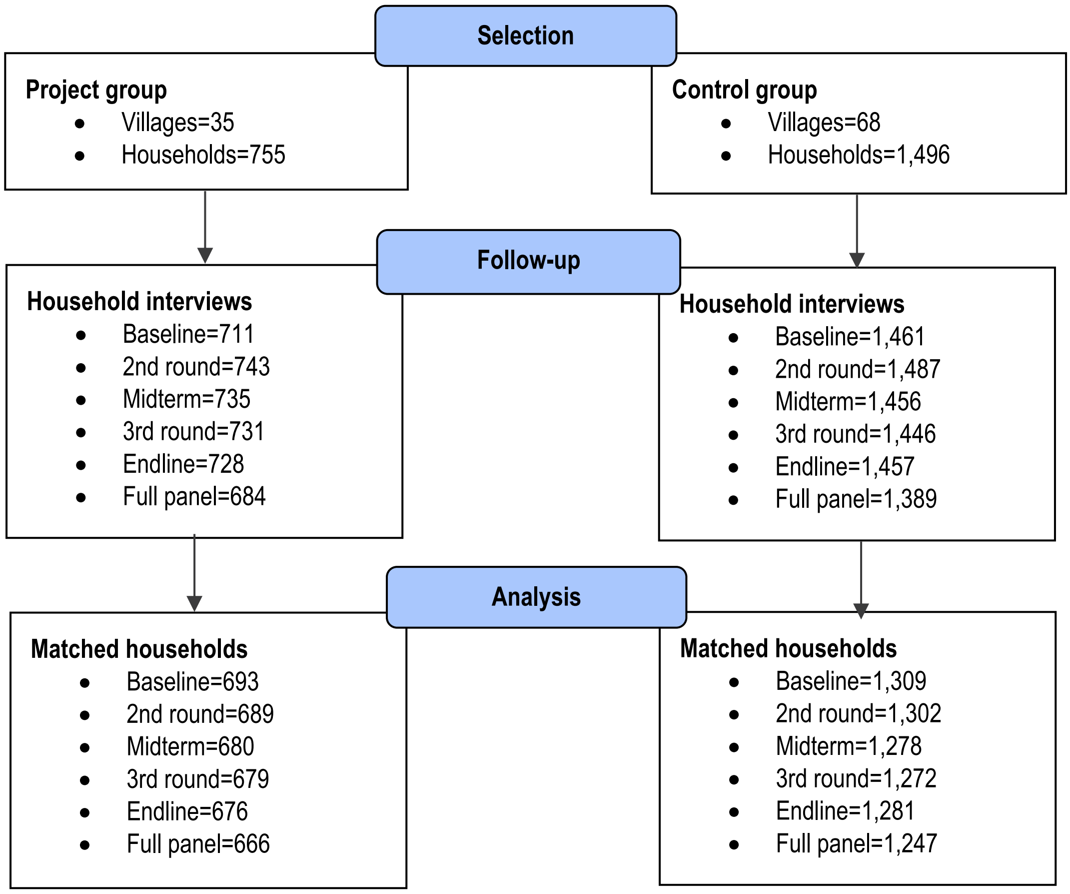
The number of interviews conducted in project and comparison villages in all rounds is shown in Table 1. The largest number of interviews was conducted in the second round, while the smallest number of interviews occurred at the baseline. There is no obvious pattern in these numbers. There are no differences in the percentage of households interviewed in MV and CV areas, suggesting that the absence of project benefits did not act as a deterrent to responding at the interviews in the CV areas. On the contrary, the numbers suggests that a proportionally larger number of households was interviewed in CV areas than in MV areas. Similarly, the proportion of households interviewed is fairly similar (and extremely high) at each round, suggesting that no dissatisfaction with the survey built up over time in either MV or CV areas.

Table 1 Household interviews in MV and CV areas

Sample	Selected	2012	2013	2014	2015	2016
MV interviews	755	711	743	735	731	728
Percentage	100%	94%	98%	97%	97%	96%
CV interviews	1,496	1,461	1,487	1,456	1,446	1,457
Percentage	100%	97%	99%	97%	97%	97%
All interviews	2,251	2,172	2,230	2,191	2,177	2,185
Percentage	100%	96%	99%	97%	97%	97%

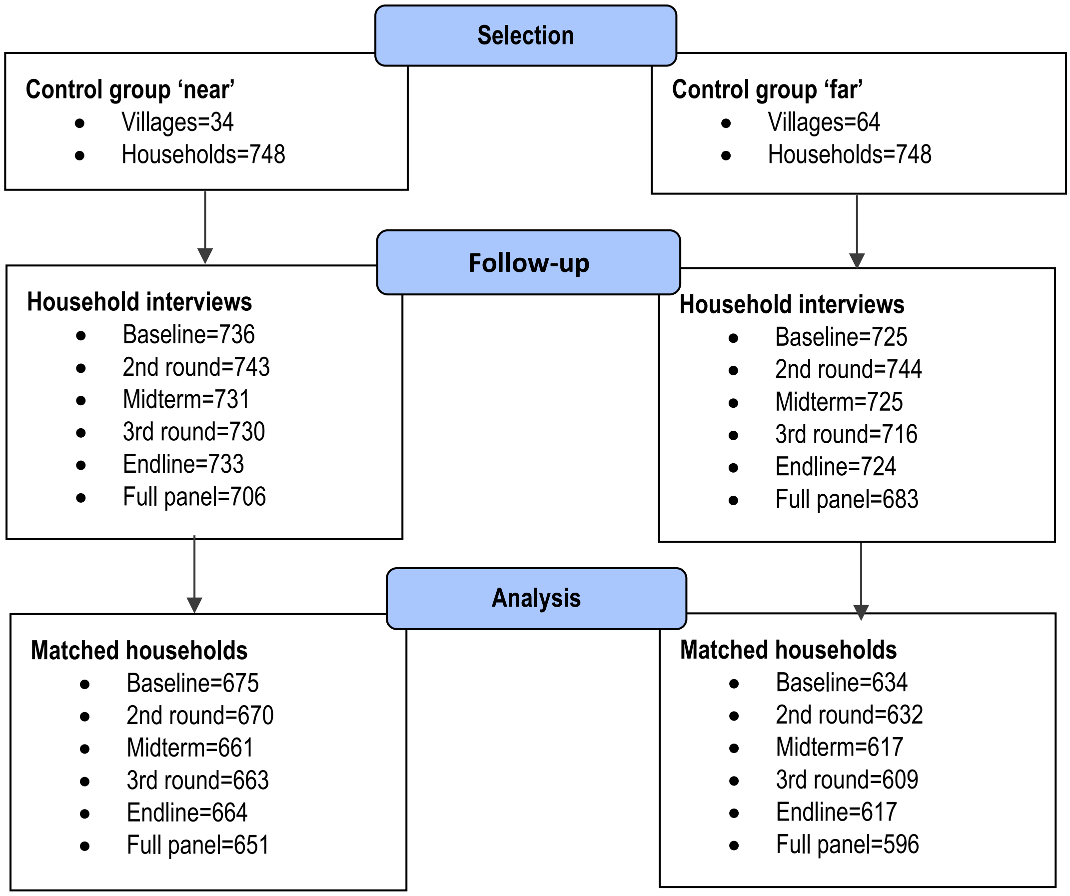
The number of households interviewed in each round is again reported in the flow diagram of Figure 3. The diagram follows the style of CONSORT diagrams commonly used in reporting the results of RCTs. The diagram also includes the number of households that were included at the analysis stage. The sample of households included at the analysis stage is smaller because it is restricted to households interviewed at baseline that were interviewed again in all the following rounds. The reason for restricting the sample to households interviewed at baseline is the need to match households, and control the estimation of project effects, using baseline characteristics. The sample of households used at the analysis stage is smaller also because of trimming of households with non-overlapping propensity scores.

Figure 3 Flow diagram of MV and CV households included in the study



For completeness we also report a flow diagram of households from the 'near' and 'far' comparison groups. By design, near and far control communities represent two separate samples. The two samples are very similar in size, though fewer households were followed up over time in the "far" comparison areas. A smaller number of "far" comparison households is also included in the final analysis because trimming of the sample following matching removed a larger number of households distant from the MV area. This is in accord with expectations as we would think that households become more and more different in characteristics as we move away from the project area.

Figure 4 Flow diagram of near and far CV households included in the study



3.2. Attrition

Studies interviewing the same households every year for several years often lose out observations over time. For example, households may lose motivation to sit the interviews or may migrate from the study area. This is unlikely to happen 'at random', and households with specific characteristics may drop out of the sample. If this happens, the results of the evaluation are valid only for the population affected by the intervention that was observed by the study and do not generalise to the wider population. Even more troublesome is the case in which households are dropping out of the sample at a different rate in the project and comparison areas. Again this is unlikely to happen 'at random' and in this case the estimated impacts not only cannot be generalised but also are biased. Monitoring the extent and the characteristics of attrition is therefore extremely important.

Table 2 Panel households in MV and CV areas

Sample	Selected	2012	2013	2014	2015	2016
MV panel interviews	755	711	707	697	689	684
Percentage		9.42%	93.6%	92.3%	91.3%	90.6%
CV panel interviews	1,496	1,461	1,454	1,424	1,391	1,389
Percentage		97.7%	97.2%	95.2%	93.0%	92.8%
All panel interviews	2,251	2,172	2,161	2,121	2,080	2,073
Percentage		96.5%	96.0%	94.2%	92.4%	92.1%

Attrition rates in the study area were very low. Less than 8% of the original target sample was lost over time. Note, however, that the original target sample refers to households selected for the survey and does not refer to the actual households interviewed at baseline, which were fewer than hoped. The attrition rate with respect to the households interviewed at baseline is only 4.6% for the whole sample ((2073-2172)/2172). More importantly, the attrition rates are very similar in MV and CV areas. They are 3.8% and 4.9%, respectively. The low level of attrition is reassuring in that our results are largely representative of the population and the estimated project effects are unlikely to be biased.

Table 3 Panel individuals in MV and CV areas

Sample	2012	2013	2014	2015	2016
MV individuals	5,231	5,576	5,854	6,021	6,338
MV panel		4,930	4,654	4,550	4,474
Percentage		94.2%	89.0%	87.0%	85.5%
CV individuals	10,337	10,649	11,023	11,255	11,750
CV panel		9,869	9,378	9,072	8,875
Percentage		95.5%	90.7%	87.8%	85.9%
Al individuals	15,568	16,225	16,877	17,276	18,088
All panels		14,799	14,032	13,622	13,349
Percentage		95.1%	90.1%	87.5%	85.7%

Attrition rates were relatively small also among individual household members (see Table 3). More than 85% of the individuals originally selected for the interview were enumerated in the last survey round. The attrition among individuals is more the result of changes in household composition and errors in reporting household membership than of actual dropping out of the study.

Table 4 Reasons for not completing the interviews

Reason	2012	2013	2014	2015	2016
No competent household member at home	21	1	8	13	12
Entire household absent	22		11	20	14
Interview postponed	10				
Interview refused	1				
Partly completed					
Dwelling vacant or destroyed		4	2	20	18
Dwelling not found	19	9	13	5	22
Household has relocated		6	15	8	
Household dissolved or deceased		1	6		
Other	6		4	7	
All	79	21	59	73	66

The number of households lost at each round and the reason for not completing the interview at each round are reported in Table 4. These numbers are very small and do not allow for investigation of whether the 'attriter' households are different from the rest of the sample. Similarly, given the small numbers, it is impossible to say whether there are differences in characteristics between attriters of MV and CV areas. The reasons for not completing the interviews are different at each survey year, but absence of a competent household member at the time of the visit, absence of the entire household and inability to find the dwelling predominate, among other explanations.

Enumerators were instructed to enquire the whereabouts of the households from neighbours in those cases in which it was found that the household had relocated. Few households appeared to have relocated and, oddly, not a single household had reportedly relocated in the year before the last survey round. Favourite locations for relocation appear to be Kumasi and surrounding areas and Accra (see Table 5).

Table 5 Whereabouts of 'relocated' households

Location	2013	2014	2015	2016
Kumasi, Kumasi, Ashanti	1	2	4	
Jagsi, Kumasi, Ashanti			1	
Delaasa, Kumasi, Ashanti			1	
Ejisu, Ejisu-Juaben, Ashanti			1	
Sariba, Northern, West Mamprusi	1			
Obuasi, Obuasi, Ashanti	2			
Luisa, Builsa, Upper East	1			
Accra, Greater Accra		7		
Kentasi, Ashanti		1		
Presetia	1			
Eastern Region			1	
Missing	0	5	0	
All	6	15	8	

3.3. Differences in trends

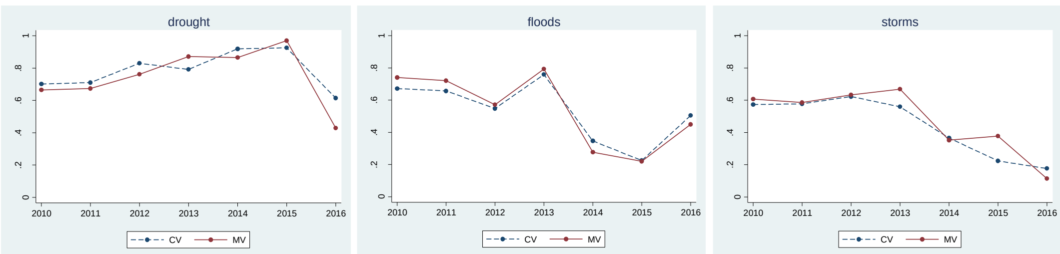
The validity of the DD approach rests on the similarity between the project and control observations in the baseline levels of the covariates and in the pre-baseline trends of the outcome variables and of the determinants of the outcomes. The baseline report tested differences in outcome variables extensively and found more than 20% of difference at the 10% statistical significance, which is larger than what we should expect by chance alone. This difference led to the decision to conduct further matching at the household level before assessing project effects.

The baseline report also investigated differences in the trends of pre-baseline outcomes. The DD 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 in the absence of 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 these cases, DD analysis may find an impact when there is none, as well as not finding an impact when there is one. The baseline report analysed differences in trends (up to two years before the baseline) in employment income, enterprise incomes and animal stock and could not find relevant differences.

In this section we extend the trend analysis conducted in the baseline report to the occurrence of agricultural shocks. In the study communities, most economics and health outcomes are influenced by weather patterns and weather shocks. The baseline survey collected retrospectively information on the occurrence of three types of shocks (droughts, floods and storms) up to two years before the baseline. In the baseline interviews households reported whether they had been significantly affected by these shocks and continued to do so in the following survey rounds.

The percentages of households reporting being affected by shocks over the previous 12 months at each survey round in MV and CV areas are presented in the charts of Figure 5 for the period from two years before the survey to the endline. Visual inspection of the charts suggests similar trends in the shocks both before and after the MV intervention.

Figure 5 Trends in weather shocks



In order to analyse more closely the differences in the occurrence of shocks between MV and CV areas we regress the occurrence of a shock against year dummies and the interaction of the year dummies with a dummy variable for MV area (in addition to a dummy for the Builsa district):

$$shock = a_1 + a_2 year_2 + a_3 year_3 + \cdots + a_7 year_7 + b_1 MV + b_2 year_2 MV + b_3 year_3 MV + \cdots + b_7 year_7 MV + cBuilsa + e$$

We then perform the following tests, jointly estimating the equality of the parameters:

- The presence of a time pattern in CV areas: $a_2 = a_3 = a_4 = a_5 = a_6 = a_7$
- The presence of a time pattern in MV areas: $a_2 + b_2 = a_3 + b_3 = a_4 + b_4 = a_5 + b_5 = a_6 + b_6 = a_7 + b_7$
- The presence of a level difference between MV and CV areas: $b_1 = b_2 = b_3 = b_4 = b_5 = b_6 = b_7 = 0$
- The presence of differences in the patterns in MV and CV areas: $a_2 = b_2,\ a_3 = b_3,\ a_4 = b_4,\ a_5 = b_5,\ a_6 = b_6,\ a_7 = b_7$

Table 6 Differences in weather shocks between MV and CV areas

	Time pattern in CV areas	Time pattern in MV areas	Level difference between MV and CV areas	Pattern difference between MV and CV areas
Droughts	77.7***	51.9***	4.2***	8.1***
Floods	51.7***	134.7***	1.2	7.6***
Storms	23.0***	39.6***	4.7***	6.0***

Note: F-test of joint hypothesis tests reported. Stars represent statistical significance levels, whereby * is P value<0.10, ** is P value<0.05 and *** is P value<0.001

Despite the similarity in trends, there appear to be different patterns in the occurrence of shocks in MV and CV areas (last column of Table 6). Year-on-year changes in the occurrence of shocks are not always in the same direction or do not always have the same intensity in MV and CV areas. In order to account for these differences, we will include year-specific shock variables in all the regression models estimating project effects.

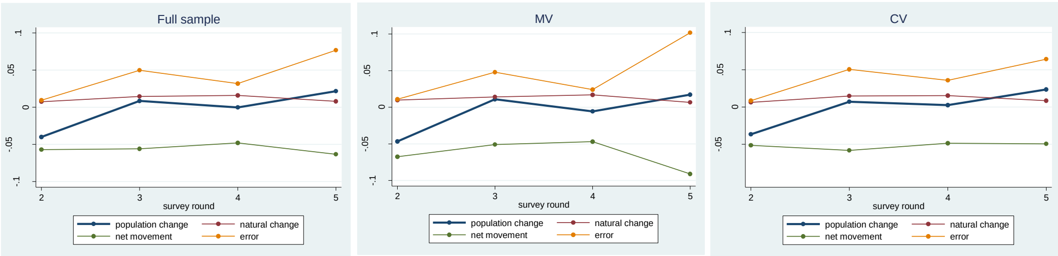
3.4. Changes in household composition

A further threat to validity of the comparison of outcomes in MV and CV areas is posed by changes in the composition of the households sampled in the MV and CV areas. For example, comparisons would be biased if a considerable portion of the CV population were to migrate to MV areas to access project benefits. In this section we analyse population changes in the study areas and particularly changes in household size.

A population can change in two ways: by natural increase (difference between births and deaths) or by movements of individuals in and out of households. In order to calculate the rate of growth in the population, and its determinants, we calculated population changes for a panel of households observed across the five survey rounds. For the purpose of this analysis, all enumerated individuals were considered households members, thus including individuals living in the household for less than six months.[5] Changes are presented in the charts of Figure 6 and reported in detail in Table 7, as percent changes with respect to the enumeration in the previous survey round. Any observed change in population that is not accounted for by natural change or by movement in and out of the household was defined as 'error'. The error category includes errors made by enumerators and respondents but is also a reflection of the somewhat fluid of household definition, whereby some individuals are considered members in some years but not in others.

[5] Household members residing in a household for fewer than six months before a survey are usually excluded from the analysis because they do not contribute significantly to household expenditure or income generation.

Figure 6 Population changes in MV and CV across survey rounds



The surveyed population is stable across the study period. The natural rate of growth (the difference between births and deaths) is about 2%, but this is more than compensated for by a net negative growth owing to individuals moving out of households. Note that this is not necessarily geographic migration, as movement in and out of households can occur within the same village or within the same geographic area. The error in reporting the number of household members is large and it appears to increase over time, possibly because enumerators were able to identify and include more and more household members over time. What is more important for our study, the changes are similar in MV and CV areas except for in the final survey year, in which there is much larger movement out of the households in MV areas.

Table 7 Population changes across surveys

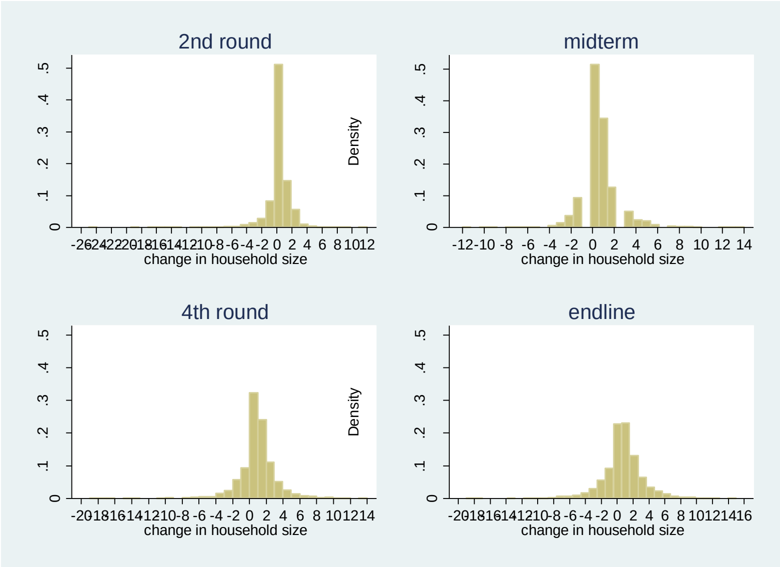
Population change	All				MV				CV			
	2013	2014	2015	2016	2013	2014	2015	2016	2013	2014	2015	2016
Overall change	-4.00	0.85	-0.01	2.15	-4.67	1.12	-0.55	1.75	-3.65	0.72	0.26	2.36
Births	1.62	2.43	2.47	2.06	2.01	2.75	2.51	2.01	1.42	2.27	2.44	2.09
Deaths	0.87	0.97	0.88	1.26	1.02	1.32	0.80	1.36	0.79	0.79	0.92	1.21
Natural change (births-deaths)	0.75	1.47	1.59	0.80	0.98	1.42	1.72	0.66	0.64	1.49	1.53	0.87
In-migration	1.26	2.31	1.95	1.74	1.44	2.89	2.25	1.99	1.17	2.02	1.80	1.61
Out-migration	6.95	7.90	6.75	8.06	8.19	7.99	6.94	11.09	6.32	7.85	6.65	6.54
Net migration (in-migrants-out-migrants)	-5.69	-5.58	-4.80	-6.33	-6.75	-5.10	-4.70	-9.10	-5.15	-5.83	-4.85	-4.93
Residual unexplained difference	0.94	4.97	3.19	7.68	1.10	4.79	2.43	10.18	0.86	5.06	3.58	6.41

Note: Figures reported are percent changes calculated with respect to the previous year

3.5. Changes in household size

Household size can change from one survey to another for a number of reasons, including natural change (births and deaths), movement in and out of the household (migration, child fostering, marriage, separations, formation of new households, etc.) and error in reporting by enumerators and respondents, which we found in the previous section to be large. The chances of changes in household size are particularly large in the context of our study because of the large household size (seven members on average) and the fluidity of household membership (Hill, 1986). There is indeed evidence that household size changed considerably across the surveys. The charts in Figure 7 show changes in household size at each round in comparison with the baseline household size. The size of the changes increases over time. At the second round, nearly 50% of households had exactly the same size as at the baseline, but at the endline only 25% of households had the same size as at baseline. In addition, the same size after four years does not necessarily mean the same people, as movements in and out of the household for several reasons can balance out. It is a legitimate question to ask whether the 'households' followed through our five survey rounds really represent the same decision unit or production/consumption unit. For the purpose of our study, however, we maintain the presumption that we are interviewing and analysing behaviour of the same household. In this we follow standard practice in applied research and we note that the changes in household size we observe do not differ greatly from changes observed in similar studies employing long panels of households (Halliday, 2005).

Figure 7 Changes in household size across surveys



From an evaluation perspective what really matters is whether the changes in household size differ between project and control areas. Household size is the denominator of per capita and expenditure figures, and changes in household size could change the outcomes. A project providing a wide range of benefits could change household size in a variety of ways, a fact that is often overlooked in evaluations. For example, MV aims at reducing the number of child deaths and maternal deaths. But MV also promotes family planning so that the final impact on natural population change is uncertain. The provision of other services and economic interventions is more likely to favour an increase in household size. For example, an improvement in living conditions could reduce migration and wealthier households may attract more marriages or child fostering. The latter may become more common to access education facilities as well. In general, we would expect MV to increase the average sample size household size.

Since some of the most important outcome variables, namely income and expenditure, are calculated on a per capita basis, changes in household size can bias the interpretation of project effects. In particular, the impact of MV on per capita expenditure and poverty can be underestimated if MV increases household size. The charts in Figure 8 show that indeed there is an increase in household size in MV areas in comparison with CV areas, while no change is visible when household size is converted into adult equivalents.[6] The difference between changes in household size and adult equivalent seems to suggest that much of the change in household size in MV areas occurs among the non-adult population, because adult equivalents are calculated by giving less weight to younger members.

[6] Income and expenditure figures are often measured per adult equivalent rather than on a per capita basis. This is because consumption needs of children are normally lower than those of adults, at least in deprived areas, and because some household goods can be shared among household members thus generating economies of scale in consumption (Deaton, 1997). This practice is also followed by the GSS, which adjusts people's expenditures in proportion to their nutritional requirements for a given age and sex, in such a way that children weigh less and count only as a fraction of an 'equivalent adult'. Thus, for example, a couple with a 2-year old child will have adult equivalent size 2.24. A couple with an infant, a child aged 4 and another child aged 10 will have adult equivalent size 3.36. The full adult equivalence table for children and adults of different age and sex is reported in Appendix C of the baseline report.

Figure 8 Household size and adult equivalents in MV and CV areas

We tested the DD impact of MV on household size and adult equivalent using three different model specifications. The impact of MV on household size is very small (about 10% of a member) and never statistically significant. The impact on adult equivalents is similarly small and negative, with P-values near statistical significance at 10%. In our application we estimate impacts on expenditure and incomes using adult equivalents rather than household size. Since MV appears to reduce the number of adult equivalents, the project effects could be somewhat overestimated but the changes are so small to be irrelevant to our analysis.

Table 8 Impact of MV on household size

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2014	DD impact 2015	DD impact 2016	DD average impact
Household size	7.08	0.11 (0.725)					
Cross section			-0.12 (0.157)	-0.17 (0.450)	0.24 (0.554)	0.22 (0.568)	0.13 (0.387)
Fixed effects			-0.12 (0.154)	0.16 (0.438)	0.24 (0.560)	0.19 (0.532)	0.11 (0.345)
Lagged model			-0.12 (0.125)	0.15 (0.407)	0.24 (0.524)	0.19 (0.519)	0.11 (0.345)
Adult equivalents	5.31	0.07 (0.758)					
Cross section			-0.12* (0.088)	-0.08 (0.176)	-0.15 (0.126)	-0.11 (0.193)	-0.11 (0.126)
Fixed effects			-0.12* (0.085)	-0.10 (0.156)	-0.15 (0.135)	-0.13 (0.161)	-0.12 (0.115)
Lagged model			-0.11* (0.061)	-0.09 (0.136)	-0.14* (0.087)	-0.12 (0.137)	-0.11* (0.082)

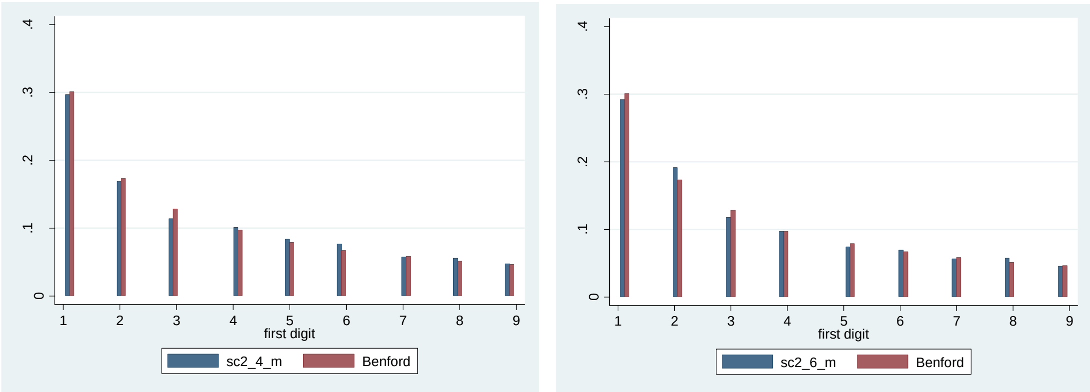
Note: Coefficient estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is P value<0.10, ** is P value<0.05 and *** is P value<0.001

3.6. Data quality

We investigate the quality of the data using an application of 'Benford's law' of first digits. The first digits of a random draw of number have an equal probability of occurring. However, this is not the case with most figures we are working with, for which the probability of being equal to 1 is about 30% rather than the expected 10%. This is in essence 'Benford's law', an empirical regularity found in a wide range of data phenomena, whereby the first digits of a series of number tend to follow a specific logarithmic function (Benford, 1938; Fewster, 2009). These regularities have been found in street numbers, bank accounts and many other applications. Because of this expected regularity in the distribution of first digits, Benford's law has been used to detect anomalies in the data or data fabrication and accounting frauds (Judge & Schechter, 2009).

To show how Benford's law works we plot in the charts of Figure 9 the frequency distribution of the first digits of reported quantity of agricultural production by a sample of Filipino farmers against the Benford's distribution. The first digits of the quantities of production reported by farmers (blue bars) follow the Benford's distribution (red bars) nicely. A chi square test of the null hypothesis that the two distributions are identical cannot be rejected (8.14 against a 10% critical value of 13.36) and the distance between the bars is as low as 0.017 (distance is measured as Euclidean distance divided by maximum possible distance – so that it lies between 0 and 1).[7] In the chart on the right we show the value of market sales of agricultural produce. The fit with the Benford's distribution is slightly worse (distance of 0.20) but the equality of the two distributions is not rejected (chi square=6.23).[8]

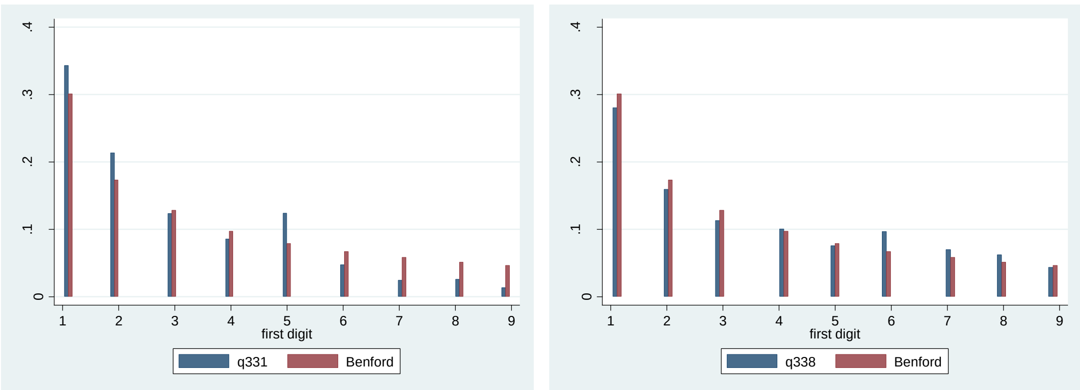
Figure 9 Quantities of agricultural production, sales and Benford's distribution (Philippines)



The charts in Figure 10 show the same distributions of first digits (agricultural production and market sales) for the MV data. The fit is good but there are some interesting differences with the Filipino charts. The digits 1, 2 and 5 appear more often than expected, particularly in the reporting of quantity produced. This suggests that some approximation of the real figure is likely to occur by either the enumerator, the respondent or both. We then continue our analysis asking the following questions:

- Are consumption figures reported with more precision than income figures?
- Are MV consumption and expenditure data worse than similar data collected in Ghana?
- Is the quality of MV data collection improving over time?

Figure 10 Quantities of agricultural production, sales and Benford's distribution (MVP)

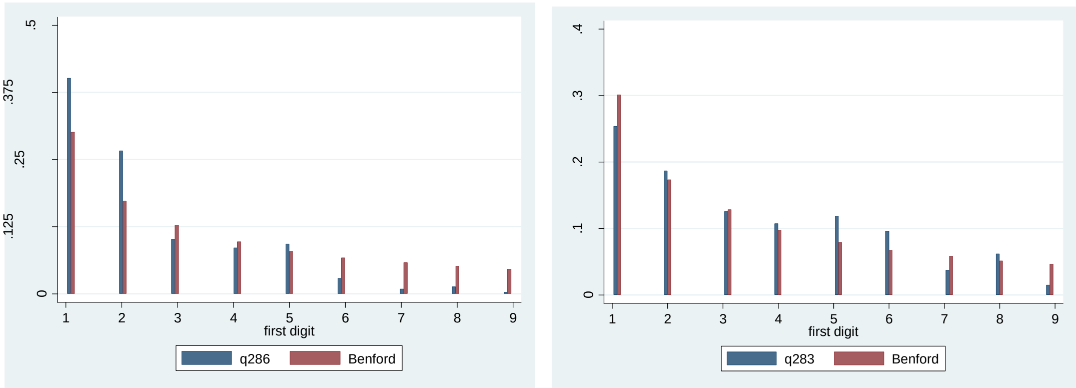


MV expenditure data appear to be less accurate than income data. The charts in Figure 11 show the first digit distributions of quantities of own consumption and values of food purchases. Perhaps a bit surprisingly, the

[7] The Euclidean distance is sqrt(SUM(observed bar-benford bar)^2), the maximum distance occurs when all first digits are equal to 9.
[8] Notice that the chi-square test is a function of the sample size, so it gets large when the sample of observations is larger. Its value should not be taken as a proxy of how 'big' is the discrepancy between the observed distribution and the Benford's distribution. This discrepancy is better measured by the Euclidean distance between the bars.

degree of approximation is larger for consumption than production figures. There is clearly an excess of first digits 1 and 2 and an absence of digits above 5 in reporting quantities of own consumption. With all the wild approximations made in measuring agricultural output, we normally think of consumption figures as being more accurate than income figures, but these charts seem to suggest that more approximation is occurring in reporting expenditures than income, something that is not limited to the MV data as will be shown shortly.

Figure 11 quantities of own consumption, food purchases and Benford's distribution (MVP)



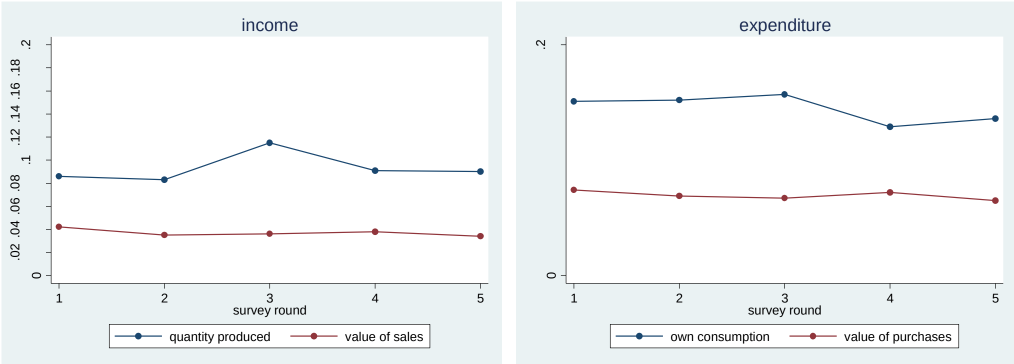
The expenditure data collected for the evaluation of MV are worse than similar data collected by household surveys in Ghana, but not much worse. We compared three different datasets: the first wave of the Yale/ISSER datasets collected in Ghana in 2011, the GLSS6 collected in Ghana in 2013-2014, and the first wave of the MV dataset, collected in the study area in 2012. The data from the ISSER and GLSS6 were restricted to rural areas of the northern regions of the country in order to make them more comparable to the MV data. Conformity to the Benford distribution is rejected for all variables considered in all datasets. From this point of view, the MV data are similar to the other datasets collected in the same area in the same period. There are, however, some differences. The distances from the theoretical Benford distribution are relatively small in the ISSER dataset, they are larger in the GLSS6 dataset and they are largest in the MV datasets. Although the MV data are of similar quality of the GLSS6 and ISSER data, they appear to be less accurate. Finally, we looked at difference in the quality of data collected in MV and CV areas (results not shown) and found none.

Table 9 MVP and GLSS6 and ISSER compared

	MVP	GLSS6	ISSER
Income data			
Quantity produced			
Distance	0.086	0.071	0.054
Chi-square	761.26	781.69	84.93
Sample size	7,509	11,476	1,807
Value of sales			
Distance	0.042	0.030	0.031
Chi-square	59.20	69.06	22.51
Sample size	2,648	4,232	585
Expenditure data			
Own-consumption			
Distance	0.151	0.112	0.092
Chi-square	4211.09	2342.37	455.74
Sample size	19,057	13,677	3,679
Value purchases			
Distance	0.074	0.133	0.069
Chi-square	2153.83	15080.54	739.98
Sample size	30,139	60,734	11,913

The quality of data collection does not appear to improve or get worse across survey rounds. We looked at distances in the distribution of the first digits using the five waves of MV survey data. The distances are on average larger for expenditure than income data. The differences are remarkably similar across rounds, suggesting that errors in reporting by enumerators or respondents are not decreasing over time.

Figure 12 Distances from Benford's distribution over time (MVP)



4. Participation in the intervention

Do people participate in the several activities promoted by the project? Who are the participants? And did the social mobilisation work lead to more social and political activity in the project communities? In this section we look at households' and individuals' participation in the various activities promoted by the intervention. We then investigate the characteristics of participants in project activities and whether they have changed over time. Finally, we look at people's involvement in the social and political life of their community and consider whether MV had any impact on collective action, trust in public services and authorities and political participation.

4.1. Participation rates

We first look at participation rates reported through household interviews for a number of project activities. We were not able to calculate rates for all project activities because at the time of designing the questionnaires the details of the interventions were not fully known. Also, some of the participation rates reported refer to interventions that were implemented with different degrees of intensity. Some of the interventions were abandoned or started late in the programme. It should also be noted that the phrasing of survey questions was not always able to capture participation in specific activities and that in the household interviews often a single respondent provides answers for all household members. For all these reasons, the reported participation rates should not be taken as exact point estimates of people's involvement in the project. However, similar biases in the calculation of participation rates occur in the control group as well and the difference between participation rates in the project and control group gives us an overall idea of the reach of the intervention.

Since at the time of study design we did not know the details of project activities, many participation rates begun to be recorded from the second round and others were recorded from the later rounds. Baseline participation rates are available for activities that were known to be implemented by MVP and other organisations in the study areas, such as, for example, visits to clinics, visits by community health workers and agricultural training. In Table 10 we report differences in participation between the project and the control group for every single round with respect to the baseline year (the first year the rates are observed) and the last column in the table shows the average participation rates over the project period. Note that similar interventions are occurring in the control areas as well and that we are reporting the difference in difference. The same rates are reported for a selected number of interventions in the charts of Figure 13.

The data clearly show that a large intervention took place across all sectors. With just two exceptions (daddies' clubs and school bursaries) all differences in participation rates between the project and the control group are statistically significant at 5%. In agriculture, there are large differences in participation in cooperatives (17.8%), farmers' groups (15.7%), the use of fertiliser (12.9%) and savings and loans associations (10.8%). Participation in social groups was particularly successful in women's groups (9.5%) and parent-teachers associations (PTAs) (17.3%). In education, there are large differences in children receiving school supplies (10.8%) and school meals

(12.9%). The most dramatic differences are found in visits by a community health worker (CHW), which are well above 40% between the project and the control group. Other health interventions were less successful but still noticeable: membership of the National Health Insurance Scheme (NHIS) (11.8%), deworming (11.8%), vitamin A (5.7%), visits to health facilities (5.5%) and distribution of bednets (8.7%).

Table 10 Households' participation in project activities

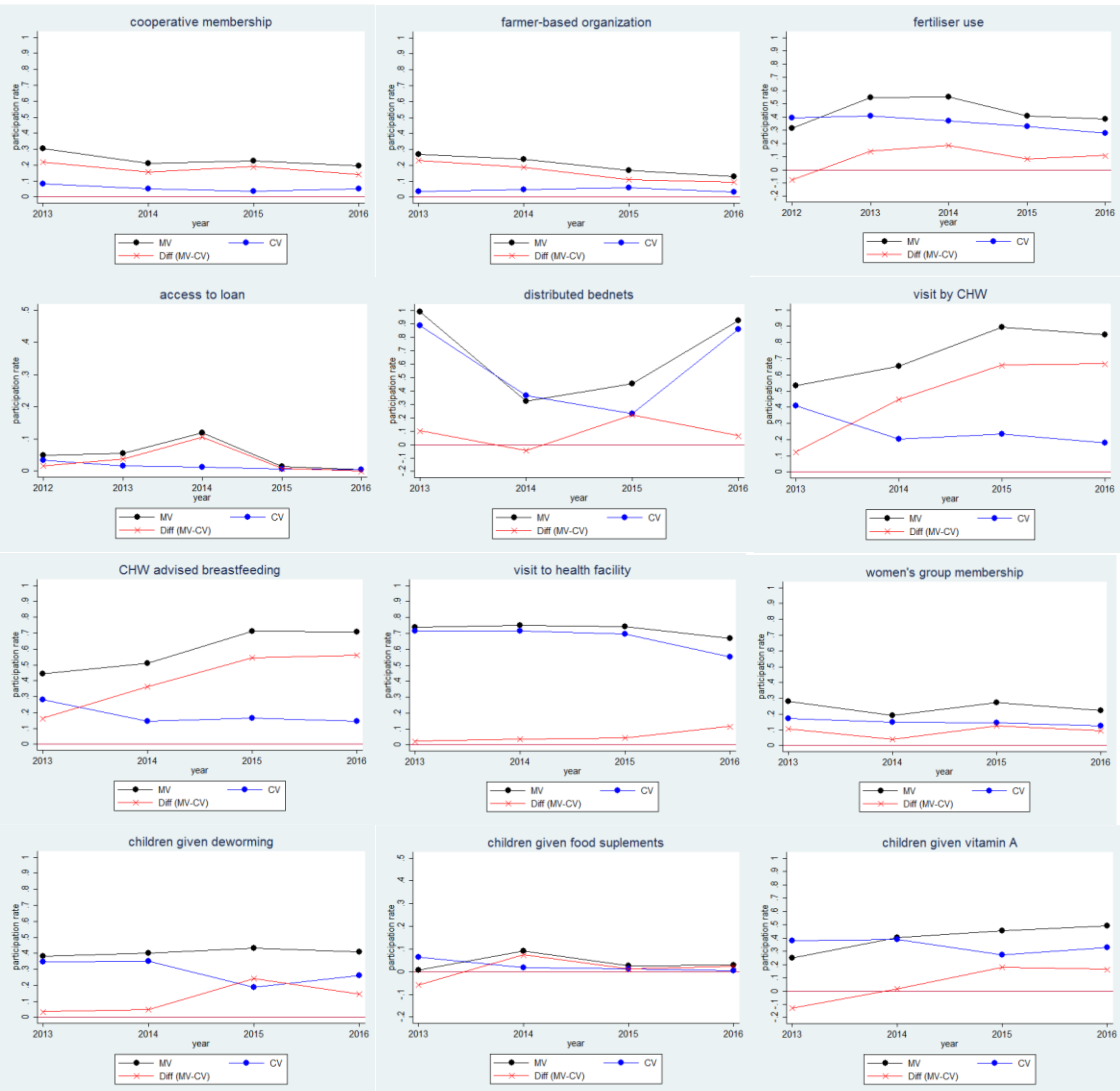
	Baseline CV	Baseline diff. MV	Comp. change 2013	Comp. change 2014	Comp. change 2015	Comp. change 2016	Average comp. change
Cooperative			21.9*** (0.000)	15.9*** (0.000)	19.1*** (0.000)	14.2*** (0.000)	17.8*** (0.000)
Farmer-based organisation			23.1*** (0.000)	18.9*** (0.001)	10.9*** (0.000)	9.7*** (0.000)	15.7*** (0.000)
Farmer field school			2.7** (0.010)	0.4 (0.398)	1.8* (0.063)	2.5 (0.117)	1.8*** (0.001)
Women's group			10.8*** (0.001)	4.1 (0.146)	12.8*** (0.001)	9.6*** (0.004)	9.3*** (0.000)
PTA			20.5*** (0.000)	5.9 (0.218)	20.2*** (0.000)	23.4*** (0.000)	17.5*** (0.000)
WASH				0.9** (0.011)	1.0** (0.027)	0.7** (0.031)	0.8*** (0.001)
MDG school club				0.0	0.0	0.6** (0.033)	0.2** (0.033)
Water and sanitation development board				0.7 (0.116)	0.9** (0.016)	1.5** (0.047)	1.0*** (0.007)
Mother-to-mother support group				0.4 (0.238)	3.3*** (0.002)	2.8** (0.049)	2.2*** (0.001)
Daddy's club				-0.1 (0.323)	0.2 (0.322)	0.4* (0.082)	0.2* (0.083)
Village savings and loan association				0.5 (0.725)	11.7*** (0.000)	20.4*** (0.000)	10.8*** (0.000)
School management committee				0.1 (0.781)	1.2** (0.015)	0.6 (0.250)	0.6* (0.081)
Any household member received a loan	3.3	1.6 (0.260)	3.8*** (0.004)	10.6*** (0.000)	0.7 (0.190)	-0.1 (0.885)	3.8*** (0.000)
Used any fertiliser	39.4	-7.7 (0.132)	14.1*** (0.002)	18.1*** (0.000)	8.3* (0.072)	10.8** (0.045)	12.9*** (0.002)
Membership of NHIS			13.9*** (0.004)	20.9*** (0.000)	8.1* (0.093)	4.2 (0.201)	11.8*** (0.001)
Someone distributed bed nets			10.3*** (0.000)	-4.3 (0.452)	22.2*** (0.000)	6.6** (0.012)	8.6*** (0.000)
Visit by CHW			12.4*** (0.000)	44.9*** (0.000)	66.0*** (0.000)	66.7*** (0.000)	47.3*** (0.000)
CHW provided condoms			4.1*** (0.006)	18.2*** (0.000)	4.0*** (0.001)	3.5*** (0.003)	7.5*** (0.000)
CHW measured children's arms			10.6*** (0.000)	34.5*** (0.000)	58.6*** (0.000)	51.6*** (0.000)	38.7*** (0.000)
CHW advised on breastfeeding			16.3*** (0.000)	36.3*** (0.000)	54.6*** (0.000)	56.0*** (0.000)	40.7*** (0.000)
CHW advised on child feeding			16.2*** (0.000)	38.9*** (0.000)	59.7*** (0.000)	59.5*** (0.000)	43.4*** (0.000)
CHW advised on use of bed nets			17.6*** (0.000)	36.9*** (0.000)	60.0*** (0.000)	66.5*** (0.000)	45.1*** (0.000)
Visited a health facility			2.4 (0.481)	3.5 (0.303)	4.5 (0.104)	11.7*** (0.003)	5.5*** (0.007)
Children given deworming			3.5 (0.509)	5.0 (0.124)	24.2*** (0.000)	14.7*** (0.000)	11.8*** (0.000)
Children given vitamin A			-12.8*** (0.001)	1.5 (0.704)	18.1*** (0.000)	16.3*** (0.000)	5.7** (0.014)
Children given food supplements			-5.7*** (0.000)	7.5*** (0.000)	1.4** (0.015)	2.3*** (0.002)	1.3** (0.039)

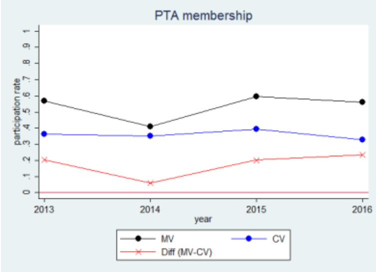
	Baseline CV	Baseline diff. MV	Comp. change 2013	Comp. change 2014	Comp. change 2015	Comp. change 2016	Average comp. change
Child had a school meal on previous day	20.1	8.9	12.0	16.3**	11.1	12.1	12.9*
		(0.218)	(0.114)	(0.027)	(0.176)	(0.135)	(0.059)
Children received a bursary			0.4	0.5	0.1	0.2	0.3
			(0.142)	(0.104)	(0.869)	(0.535)	(0.170)
Children received stationery, uniform, etc.			8.2*	4.1	12.7***	18.3***	10.8***
			(0.053)	(0.356)	(0.000)	(0.000)	(0.001)
Child given a sanitary pad			1.2*	2.7***	10.9***	5.4***	5.0***
			(0.073)	(0.006)	(0.000)	(0.003)	(0.000)

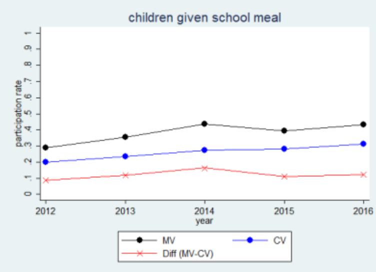
Note: Coefficients are DD estimates using a cross-sectional model, estimated using IPW method. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

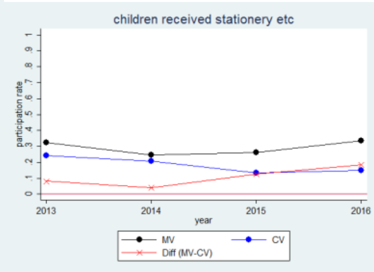
It is also interesting to look at the plots of participation rates in MV and CV areas rather than their differences. Figure 13 plots participation rates for MV and CV areas (and their differences) for a sample of the participation activities reported in Table 10. The plots show how participation rates changed in MV areas over time independently of what happened in the CV areas. In agriculture, it appears that participation in farmer-based organisations decreased over time and that a peak in the use of fertiliser and loans occurred between 2012 and 2013 (note that participation rates reported refer to 12 months before the survey interviews). Visits by CHWs increased throughout the duration of the programme, while other health interventions like the distribution of vitamin A or deworming remained relatively stable. Finally, education interventions (PTA membership, school meals and the provision of school supplies) appear to have slightly increased over the project period.

Figure 13 Participation rates in MV and CV areas









We examined participation rates also using the information emerging from the adults' surveys. The adults' questionnaires interviewed all women aged 15-49 and a sample of male respondents for each household (Table 10). The questionnaires are the same usually employed by Demographic and Health Surveys (DHS) and therefore much of the information is about health interventions. Unlike the household questionnaires, adults' questionnaires were not designed by the evaluation team and no attempt was made to capture specific project activities. However, many of the MV interventions in health and other sectors are standard and were therefore reported in the interviews.

The differences in participation rates between MV and CV are similar to those provided by the household questionnaire. All differences are statistically significant at 5%. Large differences are observed in relation to CHW visits (38.4%), visits to health facilities (11.2%), membership of farmers' groups (9.8%) and attendance of agricultural training (23.2%). Note that some of the differences between participation rates obtained from the household and the adult questionnaire depend on the different composition of the samples. The majority of adult interviews were conducted with mothers, which thus affected participation in health interventions (larger) and agricultural ones (smaller). Overall, the adult data confirm the results from the household data. MV reached a large number of households in several sectors of intervention.

Table 11 Adults' participation in project activities

	Baseline CV	Baseline diff. MV	Comp. change 2014	Comp. change 2016	Average comp. change
Group member	34.0	-11.5***	16.2***	11.0***	13.8***
		(0.002)	(0.000)	(0.000)	(0.000)
Member of CBO	3.3	-1.2	3.1***	2.4*	2.8***
		(0.141)	(0.000)	(0.095)	(0.000)
Member women's group	7.6	-3.2***	2.3*	-0.6	0.9
		(0.009)	(0.057)	(0.631)	(0.355)
Member farmers' group	7.1	0.5	11.1***	8.5***	9.8***
		(0.720)	(0.000)	(0.000)	(0.000)
Member economic organisation	3.4	-1.4*	-1.0	9.8***	4.3***
		(0.064)	(0.185)	(0.000)	(0.000)
Group leader	8.2	-2.9**	3.2***	4.9***	4.1***
		(0.010)	(0.003)	(0.000)	(0.000)
Attend any training	23.4	-4.6	31.9***	13.9***	23.2***
		(0.146)	(0.000)	(0.000)	(0.000)
Training agriculture	12.7	3.3	31.4***	7.1**	19.7***
		(0.204)	(0.000)	(0.024)	(0.000)
Training water	2.5	0.9	8.6***	-1.6	3.7**
		(0.546)	(0.000)	(0.457)	(0.030)
Training health	8.1	-3.6*	13.2***	0.8	7.3***
		(0.061)	(0.000)	(0.758)	(0.000)
Training environment	2.7	-1.3	4.4***	-0.1	2.3**
		(0.135)	(0.000)	(0.963)	(0.011)
Training business	3.1	-1.0	3.9***	2.8	3.5***
		(0.158)	(0.000)	(0.121)	(0.001)
Visited by CHW	36.8	-10.4***	28.9***	47.6***	38.4***
		(0.004)	(0.000)	(0.000)	(0.000)
Went to clinic	32.7	-3.7	14.1***	8.0***	11.2***
		(0.161)	(0.000)	(0.005)	(0.000)
Average number clinic visits	0.7	-0.1*	0.4***	0.3***	0.3***
		(0.067)	(0.000)	(0.007)	(0.000)

Note: Coefficients are DD estimates using a cross-sectional model using IPW method. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

Interviews with women of reproductive age further illustrate the reach of the health interventions through the deployment of CHWs and through higher access to health facilities. Family planning at health facilities and the use of contraceptive methods also increased significantly (Table 12).

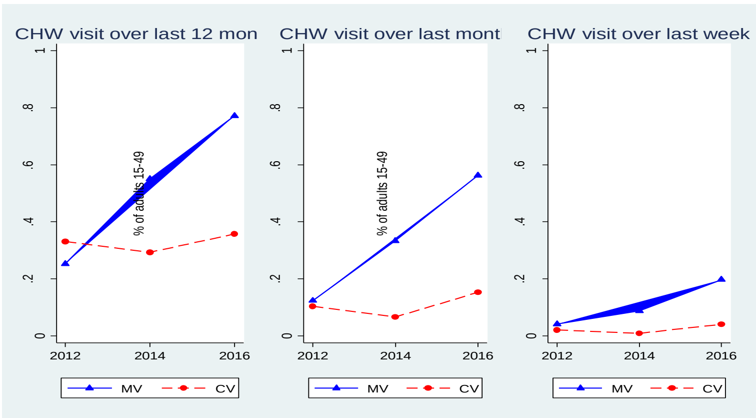
Table 12 Women's participation in project activities

	Baseline CV	Baseline diff. MV	Comp. change 2014	Comp. change 2016	Average comp. change
Using birth control method	11.7	-0.8	3.6*	11.6***	7.7***
		(0.737)	(0.062)	(0.000)	(0.000)
Visited health facility	40.5	-5.7*	21.1***	13.1***	17.5***
		(0.056)	(0.000)	(0.000)	(0.000)
Visited by CHW	29.7	-5.1*	8.4**	16.9***	12.2***
		(0.094)	(0.039)	(0.000)	(0.002)
Family planning at facility	20.4	-4.3*	12.9***	7.1**	10.2***
		(0.057)	(0.002)	(0.018)	(0.001)

Note: Coefficients are DD estimates using a cross-sectional model, using IPW method. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

We investigated CHW visits and access to health facilities a bit further given their relevance in the context of health interventions (Figure 14). By the end of the intervention, nearly 80% of all adults (male and female) reported a visit from a CHW in the previous 12 months in MV areas, 60% reported a visit by the CHW in the previous month and about 20% reported a visit in the previous week. These data confirm the reach of CHWs in MV areas, the high frequency of their visits and the increase in reach over the course of the programme.

Figure 14 Home visits by CHWs



Visits to health facilities also increased, though less spectacularly. The chart in Figure 15 illustrates visits by all adults (right chart) and female adults (left chart) to health facilities in the 12 months preceding the interviews. By the end of the programme, more than 40% of adults of both sexes reported a visit to a health facility during the previous 12 months, while 60% of mothers reported a visit. The larger attendance of health facilities by mothers is an outcome of the provision of maternal and child care by the project.

Figure 15 Adults' visits to health facilities

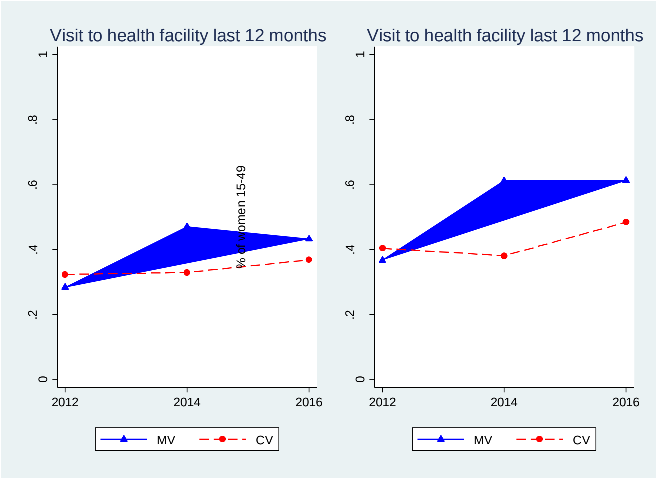


Table 13 reports the difference-in-differences from the charts of Figure 14 and 15 and performs statistical tests. The difference shown in the charts appears to be very large and highly statistically significant.

Table 13 Impact of MV on adults' visits to health facilities

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Visited by a CHW during past 12 months	33.06	-7.94**	33.02***	49.48***	40.93***
		(0.042)	(0.000)	(0.000)	(0.000)
Visited by a CHW during past month	10.45	1.91	24.22***	39.08***	31.36***
		(0.548)	(0.000)	(0.000)	(0.000)
Visited by a CHW during past week	2.31	2.01	6.05**	13.78***	9.76***
		(0.201)	(0.023)	(0.000)	(0.000)
Visited an health facility during past 12 months (all adults)	40.42	-4.92	26.37***	17.84***	22.23***
		(0.200)	(0.000)	(0.000)	(0.000)
Visited an health facility during past 12 months (women)	32.37	-3.99	17.92***	10.63**	14.40***
		(0.144)	(0.000)	(0.002)	(0.000)

Note: Coefficients are DD estimates using a cross-sectional model, using IPW method. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

Participation in maternal and child care interventions is more mixed. Table 14 reports maternal and child care services differences between MV and CV areas with respect to the baseline. The difference is only calculated with respect to the baseline because the interviews asked mothers about use of services for each child over the previous five years. Changes in service provision are more modest. There is a large increase in deliveries in health facilities and post-natal care including of newborns (post-natal checks and child weighing). The differences in access to ante-natal care services (visits by CHWs and nurses, iron tablets, malaria prevention) are much smaller, though it should be noted that the provision of these services was already very high at baseline and that the room for improvement was rather limited.

Table 14 Mothers' participation in project activities

	Baseline CV	Baseline diff. MV	Comp. change 2016
Ante-natal visit	93.0	5.2***	1.7
		(0.001)	(0.119)
Ante-natal visit with CHW	7.0	-1.5	8.8**
		(0.537)	(0.010)
Ante-natal visit with nurse	87.9	3.3	3.2*
		(0.162)	(0.057)
Iron tablets	73.9	-10.2	-5.9*
		(0.189)	(0.069)
Malaria prevention	79.9	10.4***	1.3
		(0.002)	(0.466)
HIV test	66.6	6.8**	-0.7
		(0.035)	(0.814)
Child weight	3.5	0.1	0.2***
		(0.291)	(0.000)
Child weighted after birth	23.7	-5.1	16.7***
		(0.199)	(0.003)
Child weight (average)	3.4	-0.3***	0.05
		(0.001)	(0.351)
Delivery assisted either by doctor or nurse	2.6	-0.2	4.4
		(0.917)	(0.280)
Delivery either at hospital, health centre, clinic health post	23.3	0.2	22.7***
		(0.964)	(0.000)
Check in facility	92.6	-6.2	-0.6
		(0.127)	(0.462)
Check after discharge	32.1	-13.6**	33.8***
		(0.012)	(0.000)
Breastfed	94.8	-1.8	1.8**
		(0.462)	(0.036)
Breastfed within 1 hour of birth	34.0	12.1**	5.4*
		(0.029)	(0.057)
Average months breastfed	20.3	1.2	1.7***
		(0.429)	(0.001)

Note: Coefficients are DD estimates using a cross-sectional model, using IPW method. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

Finally, the adult questionnaire collects data on children vaccinations. The data show that vaccination coverage increased in MV areas more than in CV areas but not by a large amount (Table 15). The coverage of some vaccines was already high at baseline, thus leaving little scope for improvement.

Table 15 Children's vaccinations

	Baseline CV	Baseline diff. MV	Comp. change 2014	Comp. change 2016	Average comp. change
Vaccination card	65.7	12.1***	10.0***	7.8***	9.2***
		(0.001)	(0.000)	(0.000)	(0.000)
BCG	81.8	3.2	5.1***	2.7*	4.0***
		(0.236)	(0.004)	(0.054)	(0.004)
Polio	43.3	-2.6	-2.8	-4.4	-3.4
		(0.553)	(0.451)	(0.354)	(0.365)
DPT	66.5	5.5	8.1***	5.8**	7.1***
		(0.244)	(0.008)	(0.017)	(0.003)
Measles	69.9	1.1	4.9*	5.0**	5.1**
		(0.722)	(0.051)	(0.027)	(0.010)

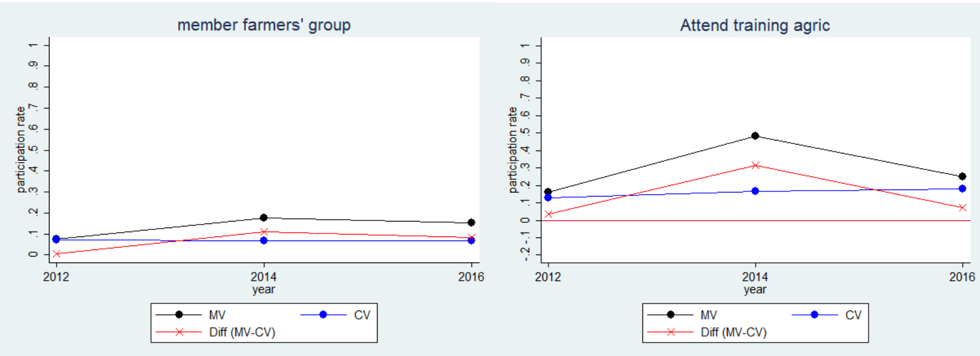
Note: Coefficients are DD estimates using a cross-sectional model, using IPW method. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

4.2. Determinants of participation in MV and CV areas

In this section we look at the characteristics of participants in the interventions. The goal of this analysis is to see whether the interventions are targeted to particular groups like, for example, women or the poor, or to see whether the intervention attracts specific socio-economic groups, like for example, large farmers or the more educated. To carry out this analysis we use the data from the adult questionnaire, which were collected at baseline, midterm and endline, and we conduct two types of comparisons. In the first comparison we show whether typical project participants are different in MV and CV areas. In the second comparison we show whether the characteristics of the participants in MV have changed as a result of the intervention, for example because the project made an extra effort to reach the poor. We conduct this analysis for four key variables from agricultural and health interventions: membership of farmers' groups, attendance of agricultural training, visits to health facilities and visits by CHWs. We find that 1) age, gender and marital status are the main factors correlated with participation in the selected activities; 2) the determinants of participation are the same in MV and CV areas before the intervention; 3) individuals of different education level, wealth and land endowments do not have any preferential access to MV interventions; and 4) the MV interventions are mildly targeted to obvious beneficiaries like mothers and farmers.

Before discussing these results in more detail we illustrate participation rates over time in the two agricultural activities selected (Figure 16). The charts in Figure 16 show that membership of farmers' groups and agricultural trainings have both increased in the MV areas and that the increase has been more pronounced at the midterm of the programme.

Figure 16 Membership of farmers' groups and agricultural training



The main determinants of participation before the intervention in both MV and CV areas are age, sex and marital status (Table 16). Members of farmers' group and those attending agricultural training are predominantly older, married and male. Wealth, education, land and other factors are irrelevant meaning that the interventions are not targeted to specific groups. After the intervention, the determinants in MV areas are somewhat different, particularly for farmers' groups. MV appears to attract (or to target) participants of older age, male, married and from households with a more educated head.

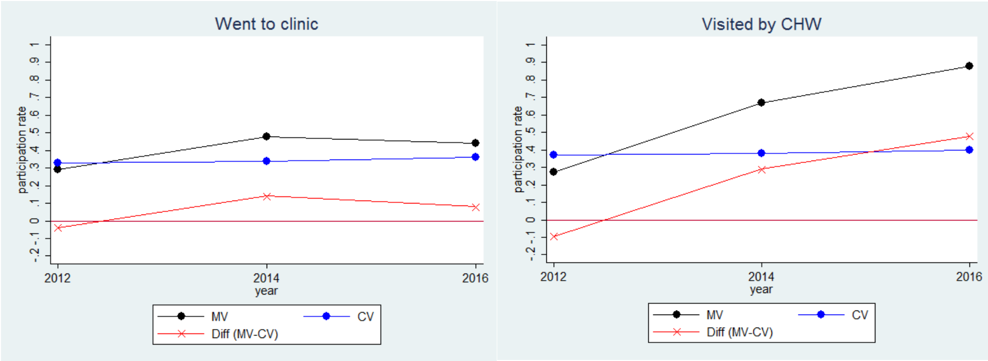
Table 16 Determinants of participation in farmers' groups and agricultural training

	Member of farmers' group			Attending agricultural training		
	Marginal effects baseline CV	Marginal effects baseline MV	Marginal effects after baseline MV	Marginal effects baseline CV	Marginal effects baseline MV	Marginal effects after baseline MV
Age	0.002***	0.001**	0.006***	0.004***	0.003**	0.006***
	(0.000)	(0.043)	(0.000)	(0.000)	(0.029)	(0.000)
Female	-0.098***	-0.081***	-0.219***	-0.080***	-0.110***	-0.194***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.051**	0.076**	0.106***	0.034**	0.089**	0.177***
	(0.002)	(0.004)	(0.000)	(0.030)	(0.009)	(0.000)
Literate	-0.019	-0.033	-0.040**	-0.060**	0.009	0.057*
	(0.360)	(0.245)	(0.088)	(0.001)	(0.830)	(0.067)
Christian	-0.003	-0.013	-0.014	-0.002	0.017	0.009
	(0.843)	(0.438)	(0.432)	(0.920)	(0.622)	(0.700)
Muslim	-0.009	-0.030	-0.012	-0.027	0.018	-0.050*
	(0.579)	(0.190)	(0.683)	(0.363)	(0.697)	(0.097)
Female-headed household	-0.014	0.001	0.058*	-0.040	0.008	-0.044
	(0.422)	(0.961)	(0.079)	(0.151)	(0.886)	(0.294)
Polygamous	-0.003	0.003	-0.005	-0.017	-0.007	-0.024
	(0.780)	(0.806)	(0.702)	(0.166)	(0.765)	(0.209)
Education 2	0.022**	0.033	0.023	0.018	0.020	0.013
	(0.030)	(0.156)	(0.220)	(0.206)	(0.547)	(0.550)
Education 3	0.020*	0.033	0.074**	0.041	0.022	0.098**
	(0.071)	(0.251)	(0.002)	(0.107)	(0.562)	(0.001)
Education 4	-0.004	0.038	0.075**	0.006	0.019	0.033
	(0.822)	(0.122)	(0.010)	(0.758)	(0.697)	(0.364)
2nd wealth quintile	-0.005	-0.022	-0.002	-0.006	0.006	-0.006
	(0.696)	(0.290)	(0.923)	(0.729)	(0.852)	(0.791)
3rd wealth quintile	-0.003	-0.026	-0.012	-0.006	0.013	-0.029
	(0.858)	(0.248)	(0.592)	(0.747)	(0.759)	(0.301)
4th wealth quintile	0.011	-0.014	0.008	0.007	0.010	0.026
	(0.482)	(0.573)	(0.720)	(0.760)	(0.823)	(0.300)
2nd land quintile	0.089	-0.015	-0.003	-0.005	-0.003	-0.017
	(0.219)	(0.445)	(0.870)	(0.806)	(0.938)	(0.516)
3rd land quintile	0.020	-0.020	0.008	0.007	0.011	-0.030
	(0.248)	(0.395)	(0.726)	(0.790)	(0.770)	(0.241)
4th land quintile	-0.002	-0.020	0.020	-0.015	0.000	-0.010
	(0.928)	(0.577)	(0.372)	(0.586)	(0.994)	(0.689)
F-test		16.17	61.24***		13.84	25.64*
		(0.512)	(0.000)		(0.678)	(0.081)
Sample	3,064	1,328	3,024	3,064	1,328	3,024

Note: The results shown are marginal effects of logit regression of participation in the activity explained by the covariates listed in the first column. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

The project resulted in a large increase in participation in the health activities considered, particularly in the proportion of people visited by CHWs (Figure 17). Interestingly, visits by CHWs increased throughout the duration of the project while visits to health facilities did not further increase in MV areas after the midterm.

Figure 17 Clinic attendance and visits by CHWs



The main characteristics associated with visits to health facilities are sex, marital status and education. People attending clinics are mostly female, married and better educated. This pattern is likely a reflection of the use of facilities for child and maternal care in both MV and CV areas. There are some minor differences between MV and CV at baseline: people visiting health facilities in MV are less likely to be female and tend to be more highly educated. After the intervention, there is a substantial increase in clinic attendance by married women, which is undoubtedly a result of the child and maternal care services offered by the project.

Before the intervention, visits by CHWs do not appear to be correlated to any specific household characteristics with two exceptions: marital status and religion. CHWs are likely to visit married families (again probably a reflection of the focus on child and maternal care) and non-Muslim families. Also, before the intervention there are no differences between MV and CV areas. After the intervention, there is visible increase of CHWs visits to married individuals and to families of Muslim religion. The latter effect probably reflects the increase in coverage by CHWs regardless of household characteristics. Also, it appears that in MV better-educated individuals are more likely to be visited.

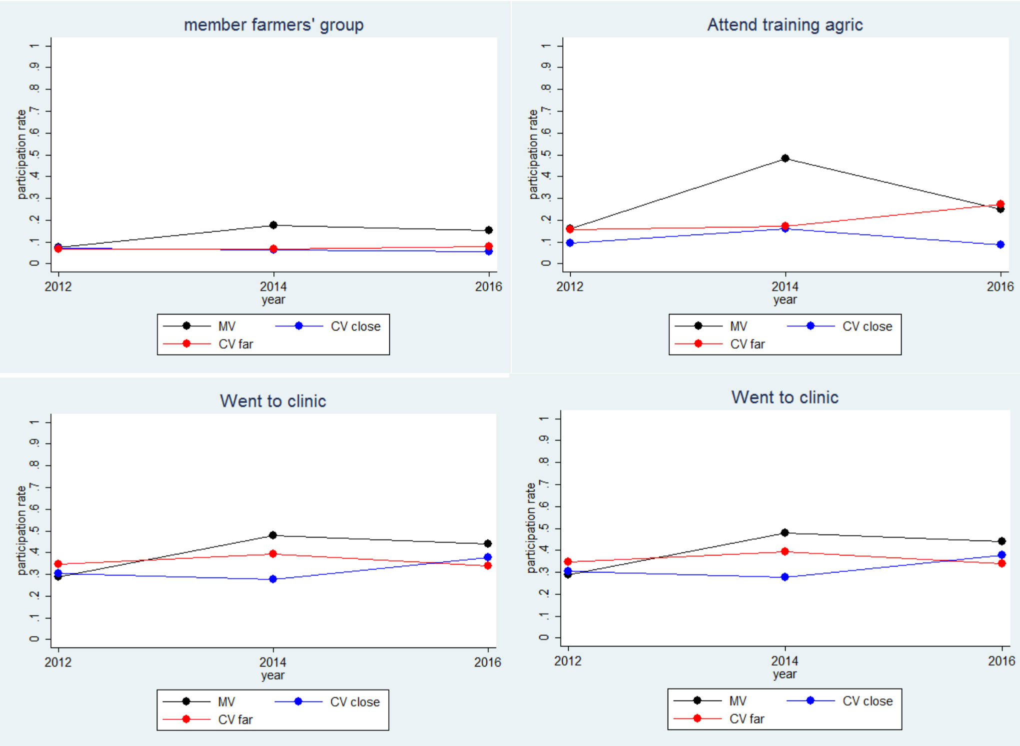
Table 17 Determinants of clinic attendance and CHW visits

	Visited a health facility			Visited by a CHW		
	Marginal effects baseline CV	Marginal effects baseline MV	Marginal effects after baseline MV	Marginal effects baseline CV	Marginal effects baseline MV	Marginal effects after baseline MV
Age	-0.001	-0.002**	-0.002	0.002*	0.000	0.001
	(0.415)	(0.038)	(0.211)	(0.098)	(0.763)	(0.222)
Female	0.115***	0.044	0.225***	-0.027	-0.009	0.017
	(0.000)	(0.122)	(0.000)	(0.287)	(0.657)	(0.250)
Married	0.135***	0.139***	0.203***	0.098***	0.088***	0.041**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)	(0.033)
Literate	0.093*	0.007	0.013	0.051	0.011	-0.003
	(0.000)	(0.889)	(0.642)	(0.139)	(0.821)	(0.926)
Christian	0.054	-0.003	0.016	-0.014	0.003	0.036
	(0.101)	(0.946)	(0.575)	(0.657)	(0.951)	(0.255)
Muslim	0.054*	-0.031	0.051	-0.074*	-0.146**	0.101**
	(0.086)	(0.407)	(0.142)	(0.051)	(0.012)	(0.013)
Female-headed household	0.047	0.003	-0.061	0.107**	-0.046	0.005
	(0.213)	(0.960)	(0.158)	(0.011)	(0.507)	(0.909)
Polygamous	-0.011	0.013	-0.008	-0.030	-0.005	0.018
	(0.523)	(0.611)	(0.582)	(0.299)	(0.881)	(0.403)
Education 2	0.036	0.095**	0.011	0.017	0.020	0.004
	(0.112)	(0.008)	(0.632)	(0.543)	(0.614)	(0.888)
Education 3	0.014	0.078**	0.034	0.023	0.027	0.069**
	(0.546)	(0.020)	(0.312)	(0.426)	(0.567)	(0.020)
Education 4	0.072**	0.103*	0.044	-0.052	-0.005	0.070*
	(0.007)	(0.055)	(0.169)	(0.203)	(0.932)	(0.098)
2nd wealth quintile	0.003	-0.012	0.020	0.079**	-0.035	-0.047
	(0.927)	(0.728)	(0.505)	(0.001)	(0.435)	(0.151)
3rd wealth quintile	0.026	0.060	0.029	0.039	-0.037	-0.032
	(0.453)	(0.169)	(0.273)	(0.179)	(0.459)	(0.323)
4th wealth quintile	0.002	0.084**	0.007	0.006	-0.028	-0.029
	(0.956)	(0.038)	(0.824)	(0.863)	(0.553)	(0.405)
2nd land quintile	0.006	0.001	-0.008	-0.012	0.014	-0.023
	(0.792)	(0.962)	(0.754)	(0.731)	(0.693)	(0.443)
3rd land quintile	-0.006	-0.054	0.015	0.067	0.044	-0.032
	(0.863)	(0.150)	(0.573)	(0.104)	(0.407)	(0.389)
4th land quintile	-0.018	-0.079*	0.038	0.063	-0.011	0.010
	(0.633)	(0.060)	(0.266)	(0.112)	(0.846)	(0.728)
F-test		32.08**	85.57***		21.85	69.85***
		(0.015)	(0.000)		(0.191)	(0.000)
Sample	3,064	1,328	3,024	3,064	1,328	3,024

Note: the results shown are marginal effects of logit regression of participation in the activity explained by the covariates listed in the first column. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

Next we investigate the presence of spill-over effects in project participation. The services offered by the project can be accessed by individuals residing outside the MV areas. This possibility has important implications for the estimation of project effects. If mothers from comparison areas are attending MV clinics, the impact of MV can be underestimated while the positive impact on people living in non-MV areas (spill-over effects) is unaccounted for. To see whether such participation spill-overs are present, we exploit the stratification of the comparison group by distance, which divides the comparison groups in 'near' and 'far' communities. Since distances between villages differ in Builsa and West Mamprusi (population and villages are much more dispersed in West Mamprusi and there are few villages at short distance) we also look at these difference separately for each district.

Figure 18 Participation rates in near and far CV areas



In principle, we would expect households from neighbouring communities to access health services more easily than agricultural extension services. It is easier to travel to a clinic in a nearby village than to join a farmer groups in another village. We would also expect spill-over effects to happen more easily in the Builsa district (where villages are closer to each other) than in West Mamprusi (where communities are dispersed across a wide geographic area). However, spill-overs can be more complicated if the district authorities or NGOs are shifting resources from one area to the other in response to the MV intervention. For example, the district authorities may have to mobilise extension agents in the MV areas thus leaving CV areas unattended.

In order to detect the presence and the direction of any spill-over effects, we estimate the ‘impact’ of MV on participation rates in project activities using only observations from the comparison group. Within the comparison group we look at the ‘impact’ of MV in near CV communities versus far CV communities and we do so separately for the districts of Builsa and West Mamprusi. The results we obtain are mixed. When comparison areas are considered altogether, the only discernible effect of MV on participation rates in ‘near’ communities is negative for visits by health workers. When disaggregating by district we notice that 1) effects are normally larger in Builsa (as expected because of different geographic configuration of the districts), but with the exception of CHWs visits; 2) participation rates in agricultural training are larger in ‘near’ communities in Builsa but smaller to the same extent in ‘near’ communities in West Mamprusi, which could be the result of different strategies in the deployment of extension agents by the two districts; and 3) CHWs visits are fewer in ‘near’ Builsa villages, which could be a result of the internal redeployment of CHWs by district authorities.

This analysis suggests that there are no obvious spill-over effects and that they do not operate in the most obvious and expected fashion. We find no evidence that participation rates in the health and agricultural activities promoted by the project are higher in communities neighbouring MV localities. We also find effects on participation rates that have opposite sign to what is expected. Some participation rates in the activities promoted by the intervention appear to decrease rather than increase in nearby communities. We speculate that this might be the result of allocation policies of CHWs and extension agents between MV and CV area conducted by the district authorities.

Table 18 Participation spill-overs

	DD impact in CVN	DD impact in CVN West Mamprusi	DD impact in CVN 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
	(1.520)	(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: marginal effects from logit model including cross-derivatives. Z-values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

4.3. Participation in social and political life

MVP invested considerable efforts in social mobilisation over a protracted period in different areas of social life through, among others, women's group, PTAs and WASH groups. It is a legitimate to ask whether the project had an impact on the quality of social and political relations. The adult questionnaires contain a number of questions, which are standard in the DHS, to get some insights into aspects of social life like collective action, trust, confidence and political participation. We estimated the impact of MV on adults' responses to a set of questions on community participation. The results are reported for every single question in Table 19 together with the results for a simple index for sets of questions on trust, confidence and political participation. The project appears to have increased people's collective action, trust and self-confidence. The impacts on political participation are more mixed and the overall impact on the index is nil.

Table 19 Impact of MV on collective action, trust and political participation

	Baseline CV	Baseline diff. MV	Comp. change 2014	Comp. change 2016	Average comp. change	Sample size
Collective action	0.277	-0.010	0.057	0.067	0.061*	12,604
		(0.774)	(0.161)	(0.111)	(0.095)	
Trust in National government	0.375	0.071	0.124	0.051	0.088	11,414
		(0.190)	(0.144)	(0.291)	(0.123)	
Trust in local government	0.360	0.028	0.180**	0.121***	0.151***	11,429
		(0.540)	(0.014)	(0.009)	(0.002)	
Trust in village leaders	0.538	-0.044	0.194**	0.165***	0.180***	11,820
		(0.403)	(0.021)	(0.006)	(0.007)	
Trust in neighbours	0.533	-0.058	0.103	0.296***	0.199**	11,869
		(0.290)	(0.278)	(0.000)	(0.013)	
Trust in doctors	0.509	-0.049	0.111	0.218***	0.166**	11,433
		(0.375)	(0.209)	(0.000)	(0.013)	
Trust in teachers	0.475	-0.006	0.076	0.160***	0.118*	11,608
		(0.899)	(0.439)	(0.004)	(0.087)	
Trust in police	0.315	0.006	0.092	0.129**	0.113*	10,699
		(0.902)	(0.288)	(0.034)	(0.077)	
Trust in judges	0.328	-0.012	0.116	0.104*	0.111*	10,136
		(0.821)	(0.202)	(0.099)	(0.077)	
Trust in groups I belong	0.480	0.041	0.145	0.281***	0.212***	9,807
		(0.502)	(0.130)	(0.000)	(0.003)	
Trust in people near villages	0.352	0.044	0.047	0.110	0.082	10,663
		(0.461)	(0.606)	(0.139)	(0.267)	
Trust index	0.438	0.022	0.091	0.139***	0.115*	12,128
		(0.640)	(0.260)	(0.004)	(0.050)	
Confidence speaking in the community	0.691	-0.036	0.092**	0.025	0.059	12,664
		(0.362)	(0.032)	(0.579)	(0.133)	
Confidence speaking to neighbours	0.769	-0.048	0.214***	0.093**	0.156***	12,672
		(0.137)	(0.000)	(0.017)	(0.000)	
Confidence speaking index	1.203	-0.181***	0.366***	0.207***	0.289***	12,697
		(0.004)	(0.000)	(0.002)	(0.000)	
Vote elections	0.693	0.010		-0.027	-0.027	8,189
		(0.641)		(0.205)	(0.205)	
Local media	0.053	-0.030**		0.053***	0.053***	8,177
		(0.017)		(0.001)	(0.001)	
Informative campaign	0.070	0.013		0.009	0.009	8,175
		(0.452)		(0.741)	(0.741)	
Electoral campaign	0.079	0.011		-0.040**	-0.040**	8,179
		(0.524)		(0.039)	(0.039)	
Demonstration	0.041	-0.030***		0.039***	0.039***	8,159
		(0.001)		(0.001)	(0.001)	
Report a problem to police	0.065	-0.022*		0.028*	0.028*	8,168
		(0.069)		(0.068)	(0.068)	
Volunteer with NGO	0.164	0.038		-0.052	-0.052	8,163
		(0.345)		(0.346)	(0.346)	
Political participation index	0.167	-0.000		-0.000	-0.000	8,189
		(0.986)		(0.988)	(0.988)	

Note: Coefficients are DD estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

The first interesting result is that the project increased the percentage of people who reported having joined efforts to solve a community problem in the 12 months before the interview (this is called 'collective action' in the adult survey). The project also increased trust in all institutions considered (village leaders, neighbours, doctors, teachers, police, judges, groups) with the exception of trust in the national government and trust in people from nearby villages. The project did not improve people's confidence in speaking in public in the community but increased confidence in providing advice to neighbours. The project had a mixed impact on some political activities. The questions on political activities refer to the four years preceding the interview and therefore we only report the comparison between the endline and the baseline. MV increased the number of people bringing issues to the attention of local media, the number of people reporting problems to the police and community leaders, and the number of people taking part in protests or demonstrations, but decreased the number of people actively participating in election campaigns.

5. MV impact on the MDGs

5.1. Impact on the MDGs

In this section we adopt a 'dashboard' approach to evaluate the impact of the intervention. In the dashboard approach all the main outcomes of an interventions are presented in a table or charts. We make no attempt to aggregate impacts by families of outcomes or by using a single index. There are a number of advantages in following this approach (Stiglitz, Sen, & Fitoussi, 2009). First, the dashboard approach enables the display of the baseline values of all the variables considered. These values were largely unknown at the time of the design of the project and of the evaluation and before conducting the baseline survey. They provide a profile of living standard in the area that can be compared to the rest of the country or to similar areas. Second, though a dashboard approach does not discuss causal mechanisms determining the outcomes, key indicators can be related to project activities to detect impacts or obvious cases of lack of impacts. Third, the approach allows the identification of areas of success and failure. For example, was the project more successful in education or in health? And did it have an impact on gender inequalities? Finally, the approach allows the detection of anomalies and conflicting results. Anomalies are unexpected impacts or the absence of impacts in areas in which the project invested heavily. Conflicting results are outcomes that somehow cancel each other out such as, for example, an increase in school attendance that happens to the expenses of employment and economic activities.

In the dashboard approach we employ the MDGs. The choice of the MDGs may seem peculiar given that countries have recently adopted a new and more comprehensive set of development goals known as the Sustainable Development Goals (SDGs). However, we investigate the impact of the intervention on the MDGs for the following reasons: 1) achievement of the MDGs is the original project goal and the project activities were chosen and designed to reach this aim. It seems natural that the impact of an intervention should be first measured against the ultimate goals it set out to achieve;[9] 2) the analysis plan of the evaluation formulated in 2013 clearly stated the objective of testing the impact of the intervention on the MDGs. A deviation from the analysis plan at this stage would be problematic for the credibility of any results shown; 3) the quantitative survey instruments were largely designed to assess changes in the MDGs rather than in the SDGs or on other indicator; and 4) finally, the quantitative impact of the project on the MDGs provides a first approximation of overall project effectiveness. At the end of this section we discuss the limitations of the assessment of the MDGs and the remainder of the report is largely dedicated to assessing the impact of MV on several other welfare dimension in an exploratory way.

Obtainment of MDG indicators according to household survey data was developed following UN instructions on the measurement of the MDGs.[10] Because of the characteristics of our survey instruments, in some cases our indicators differ slightly from the official UN definitions but great care was taken in reproducing the official methodology. Table 20 details how we built each indicator and what the indicator measures represents. The table also indicates whether the indicator is observed for the same individuals, or households, over time (panel) and also the number of observations made. Some survey instruments were administered every year, while some others were only administered at the baseline, midterm and endline.

Table 20 Description of MDG indicators

MDG	Obs.	Panel	Indicator

[9] See http://www.unmillenniumproject.org
[10] See http://mdgs.un.org/unsd/mi/wiki/MainPage.ashx

1.1 Proportion of population below $1 (PPP) per day	5	Yes	The proportion of the population below the international *poverty line* of $1.25 a day at purchasing power parity, thus adjusting for cheaper cost of living in Ghana. The UN recommendations are indifferent regarding the use of income or consumption for this indicator. We decided to use household income to introduce more information to the information already provided by the poverty headcount and the food poverty headcount that are based on household expenditure.
1.2 Proportion of population below the national poverty line	5	Yes	The proportion population living below the official Ghanaian *national poverty line* allowing the purchase of a minimum basket of food and non-food items.
1.3 Poverty gap ratio	5	Yes	The mean shortfall of population from the national *poverty line*. It measures the depth of poverty by calculating how far the poor from the poverty line are.
1.4 Share of poorest quintile in national consumption	5	Yes	The share of total expenditure in the study area that goes to the poorest 20% of the population. This is a measure of inequality in the population.
1.5 Employment to population ratio	5	No	The percentage of individuals older than 15 who did any work, paid or unpaid, over the previous year not including domestic work.
1.6 Proportion of employed people living below $1.25 income (PPP) per day	5	No	The percentage of the employed (as defined above) who are income poor. This indicator was developed to provide a measure of the lack of decent work in a country.
1.7 Proportion of own account and contributing family workers in total employment	5	No	The proportion of the employed population (as defined above), engaged in farming, animal husbandry, fishery or any other self-employment without being remunerated.
1.8 Percentage of underweight children under-5	3	No	The percentage of children aged 0-59 months, whose weight is below the WHO international benchmark. It is a composite indicator of acute (wasting) and chronic (stunting) undernutrition.
1.9 Proportion of population below minimum level of dietary energy consumption	5	Yes	The proportion of individuals below the Ghanaian official food poverty line which allows the purchase of a minimum basket of food items.
2.1 Net enrolment ratio in primary education	5	No	The proportion of children of official primary school age (6-11) that are reported having attended primary school at any time during the previous year.
2.2 Proportion of pupils starting grade 1 who reach last grade of primary	5	No	The proportion of children age 11-14 who completed primary among those who ever attended primary school.
2.3 Literacy rate of 15-24 year olds, women and men	3	No	The proportion of adults age 15-24 who were able to read correctly two English sentences ('The child is playing with the ball'; 'Farming is hard work') and to do some arithmetic (9+4 and 4x5).
3.1 Ratio of girls to boys in primary education	5	No	The ratio of the net attendance rate in primary school of boys and girls age 6-11. A ratio below one implies fewer girls are attending primary than boys.
3.2 Share of women in wage employment in the non-agricultural sector	5	No	The proportion of women above 15 in overall employment in the non-agricultural sector. The indicator measures to what extent women have equal access to jobs outside agriculture.
4.1 Under-5 mortality rate	2	No	It is the child's probability of dying before 5 years of age, calculated per thousand of population over the 5 years preceding the interview using the DHS method.
4.2 Infant mortality rate	2	No	It is the child's probability of dying before 12 months of age, calculated per thousand of population over the 5 years preceding the interview using the DHS method.
4.3 Proportion of 1-year-old children immunised against measles	3	No	The proportion of children aged 0 or 1 whose vaccination card reports a measles vaccination or whose mother recall the child being given an injection in the upper arm to prevent measles.
5.2 Proportion of births attended by skilled health personnel	3	No	The proportion of deliveries among women age 15-49 assisted either by doctor, clinical officer, or nurse for all children of age 0-2 at the time of the interview. This is equivalent to UN manual definition of % of women age 15-49 with a live birth in a given time period that received antenatal care provided by skilled health personnel.

5.3 Contraceptive prevalence rate	3	No	The proportion of women aged 15-49 using any contraceptive method at the time of the interview (sterilisation, pill, IUD, injections, implants, condoms, rhythm, abstinence and withdrawal).
5.4 Adolescent birth rate	2	No	The proportion of women aged 15-19 that gave birth during the previous 5 years.
5.5 Antenatal care coverage	3	No	The percentage of women aged 15-49 who received at least one antenatal visit (doctor, clinical officer, nurse, midwife, CHW) for children who are aged 0-2 years at the time of the interview.
6.3 Proportion of population aged 15-24 with comprehensive correct knowledge about HIV	3	No	The proportion of population aged 15-49 that answered correctly 8 (yes/no) questions about obvious causes of HIV infection transmission.
6.6 Malaria prevalence among children under 5	3	No	The proportion of children with malaria based on microscopic analysis of parasite count in the blood
6.7 Proportion of children under-5 sleeping under insecticide treated bed nets	3	No	The proportion of children aged 0-59 months who slept under an *insecticide-treated mosquito net* the night before the interview.
6.8 Proportion of children under 5 with fever who are treated with anti-malarial drugs	3	No	The proportion of children aged 0-59 months with fever in the last 2 weeks who received anti-malarial drugs.
7.8 Proportion of the population using an improved drinking water source	3	Yes	The percentage of households whose main source of drinking water is: piped into welling, yard or plot; public tap; tube well and borehole; protected dug well; protected spring; bottles; sachet water.
7.9 Proportion of the population using an improved sanitation facility	3	Yes	The percentage of households that normally uses toilets: flush to piped sewer system; flush to septic tank; flush to pit (latrine); ventilated improved pit latrine; pit latrine with slab.
8.14 Fixed telephone subscriptions for 100 inhabitants	3	Yes	Percentage of households reporting having a landline in the home.
8.15 Mobile cellular subscriptions for 100 inhabitants	3	Yes	Percentage of adults aged 15-49 reporting a personal use of a mobile phone during some or all the year before the interview.

First we compare the available MDG indicators in the MV and CV areas at the baseline. We also compare the same indicators to the values prevailing in the rest of the country (Table 21). The MDG indicators for all Ghana are those reported by the latest DHS report (Ghana Statistical Service, Ghana Health Service (GHS), & ICF International, 2015), with the exception of the poverty data that are those reported by the latest Ghana Statistical Service (GSS) poverty report available (Ghana Statistical Service, 2014b) Some interesting observations can be made by comparing the MDG indicators in the study area to the whole country. First, the area is much poorer than Ghana in terms of monetary poverty. The difference in monetary poverty is huge, though the percentage of malnourished children is not much higher. Second, primary school attendance rates are comparable to those prevailing in Ghana. However, the literacy rates among young adults (15-24 years of age) are extremely large, which suggests the study area has made very great progress in students' enrolments in recent years. Third, there is no disparity in school attendance in Ghana between boys and girls, while in the study areas girls are more likely to attend school than boys and the difference increases at higher school levels. Fourth, child mortality rates are much higher in the study areas and nearly half of children are immunised against measles. Fifth, there are some large differences in maternal and child health. The proportion of mothers receiving ante-natal care is similar and a larger number of children sleep under treated bednets. However, the proportion of births attended by skilled professionals is much smaller and there are large differences also in the use of contraceptives and knowledge of HIV. Finally, households in the study area have better access to improved water sources than the rest of the country but much less access to toilet facilities. In summary, these data suggest that the study area is much poorer in monetary terms, that child mortality rates are higher and that maternal health and health services are poorer and that access to toilet facilities is limited, while rates of school attendance, access to water facilities and undernourishment are similar.

Table 21 Baseline MDGs in MV, CV areas and Ghana

MDG	Baseline CV	Baseline diff. MV	Ghana
Proportion of population below $1 (PPP) per day	83.52	0.13 (0.515)	

MDG	Baseline CV	Baseline diff. MV	Ghana
Proportion of population below the national poverty line	88.08	-0.47 (0.836)	24.2
Poverty gap ratio	48.72	0.96 (0.699)	7.8
Share of poorest quintile in national consumption	7.24	-0.59 (0.777)	
Employment to population ratio	79.49	-3.19 (0.217)	
Proportion of employed people living below $1 (PPP) per day	52.75	8.04 (0.488)	
Proportion of own account and contributing family workers in total employment	95.86	-3.56 (0.080)	
Percentage of underweight children under-5	16.43	-1.78 (0.345)	11.0
Proportion of population below minimum level of dietary energy consumption	66.48	0.94 (0.798)	8.4
Net attendance ratio in primary education	69.56	-9.30 (0.056)	70.6
Proportion of pupils starting grade 1 who reach last grade of primary	74.54	2.86 (0.602)	
Literacy rate of 15-24 year olds, women and men	32.93	-2.52 (0.505)	85.1
Ratio of girls to boys in primary education	1.04	**0.24**** (0.006)	1.0
Share of women in wage employment in the non-agricultural sector	52.75	8.04 (0.488)	
Under-5 mortality rate	103.57	-34.66 (0.030)	60
Infant mortality rate	69.94	-25.67 (0.097)	41
Proportion of 1-year-old children immunised against measles	50.89	10.44 (0.015)	89.3
Proportion of births attended by skilled health personnel	29.95	-0.69 (0.902)	73.7
Contraceptive prevalence rate	9.77	-0.31 (0.884)	26.7
Antenatal care coverage	79.63	**5.76*** (0.074)	97.0
Proportion of population aged 15-24 with comprehensive correct knowledge about HIV	12.75	0.20 (0.920)	23.6
Proportion of children under-5 sleeping under insecticide treated bed nets	54.64	**-23.76***** (0.000)	46.6
Proportion of the population using an improved drinking water source	72.54	-0.037 (0.946)	64.2
Proportion of the population using an improved sanitation facility	8.93	0.32 (0.923)	15.0
Fixed telephone subscriptions for 100 inhabitants	0.19	-0.19 (0.194)	
Mobile cellular subscriptions for 100 inhabitants	45.73	3.18 (0.419)	

Note: P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

The data reported in Table 21 also allow for comparing baseline indicators in MV and CV areas. Comparisons are made after matching households and trimming the sample to remove households in either group that are too different from the comparator sample. Rates of poverty, malnourishment and other distributional indicators of poverty are very similar in the MV and CV areas. Fewer children are attending primary school areas at the baseline though the difference is not statistically significant. Many more girls are attending primary school in the MV area and the difference is statistically significant. Child and infant mortality rates are much smaller in MV area though the difference is not statistically significant. Ante-natal coverage is higher in MV areas while the proportion of children sleeping under bednets is much smaller, though this latter result is likely to be the result of seasonal factors related to the implementation of the baseline survey at different times of the year. Access to basic infrastructure such as water, toilets and mobile communication is very similar.

The trends in all MDGs in MV and CV areas are displayed in the charts of Figures 19 to 25. As discussed in Section 2 (on the evaluation design), unlike most evaluations, the MV evaluation adopted sampling proportional to population size, rather than sampling a fixed number of observations from each cluster. This resulted in the estimation of individual-level project effects rather than village-level project effects. If the impact of MV varies with village population size, village-level and individual-level estimates differ. In Appendix B we show that the impacts of MV are slightly larger when estimated using village-level estimates, suggesting that the interventions have larger impact in small villages than in large villages.

Figure 19 MDG 1: Eradicate extreme poverty and hunger

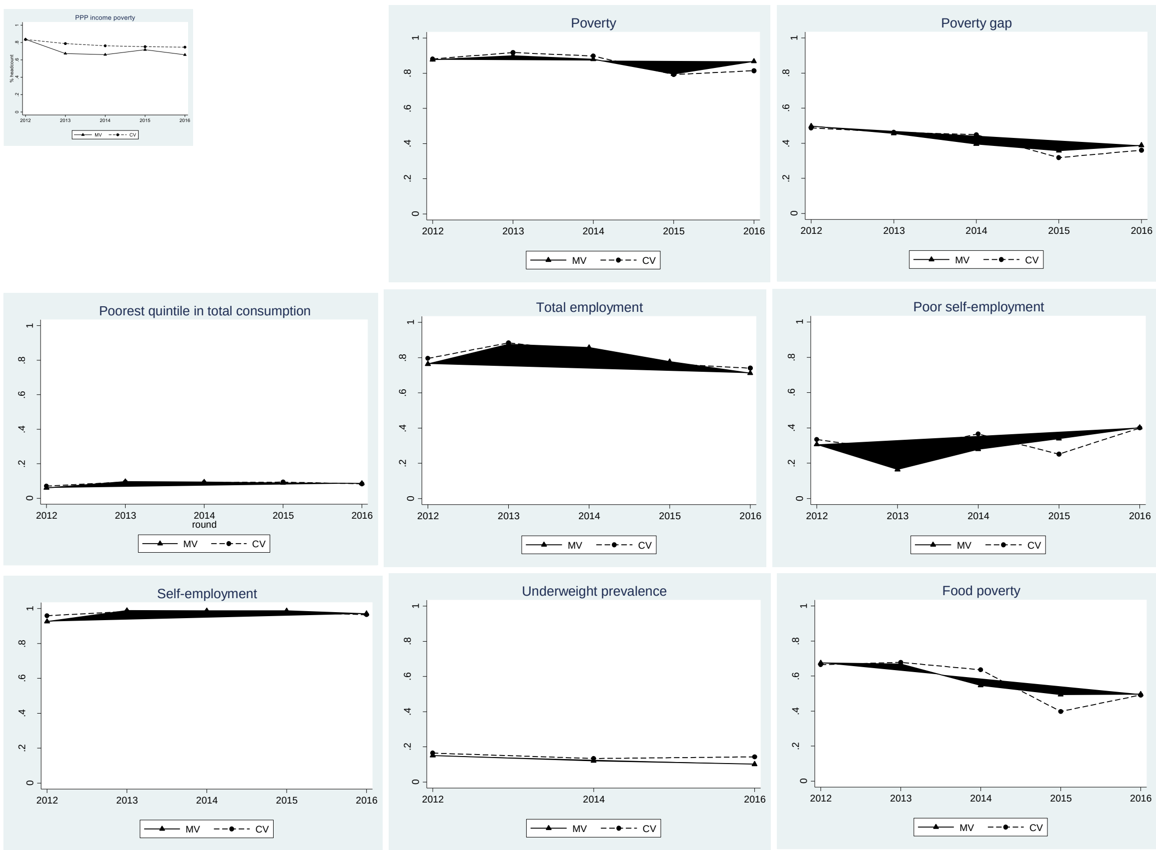


Figure 20 MDG 2: Achieve universal primary education

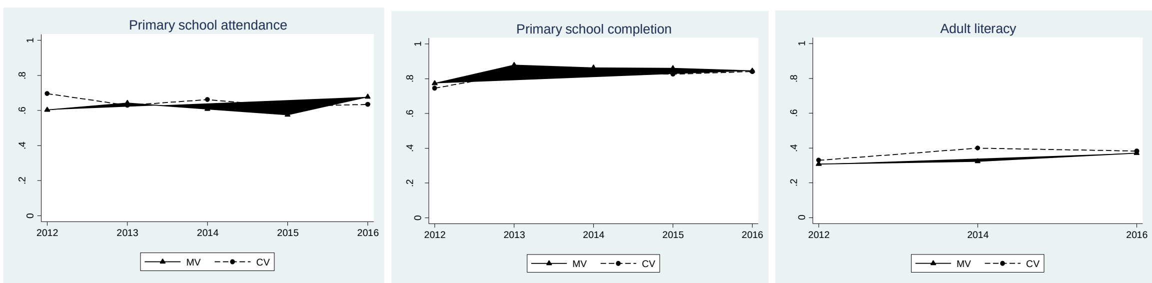


Figure 21 MDG 3: promote gender equality and empower women

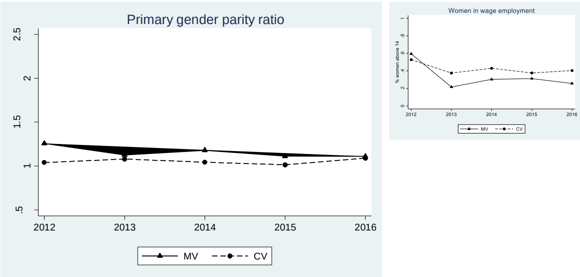


Figure 22 MDG 4: Reduce child mortality

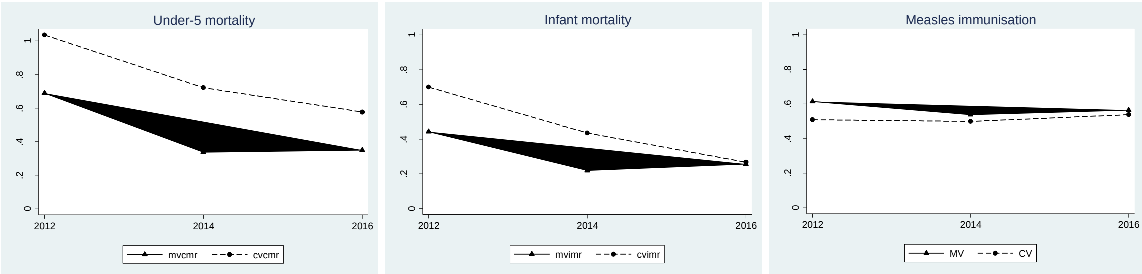


Figure 23 MDG 5: Improve maternal health

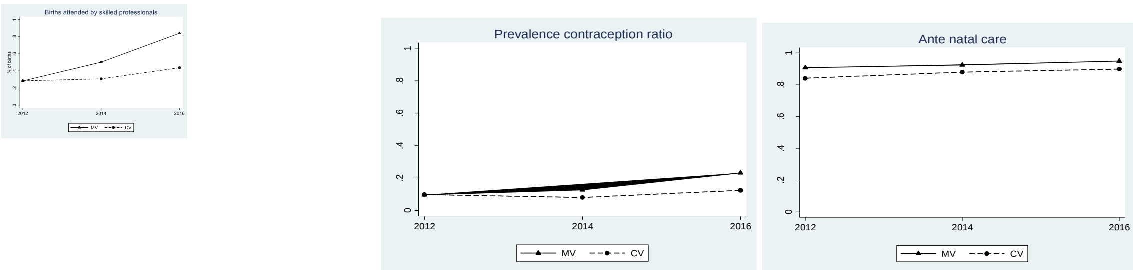


Figure 24 MDG 5: Improve maternal health

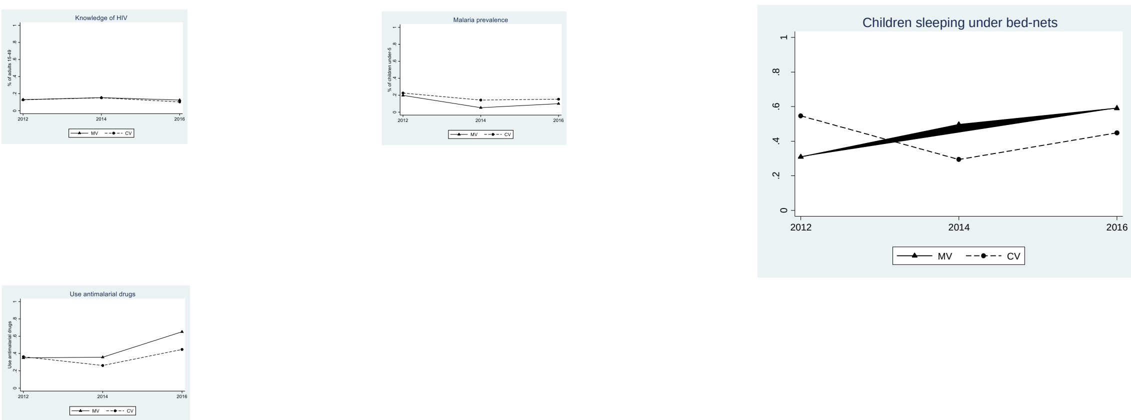


Figure 25 MDG 7: Ensure environmental sustainability

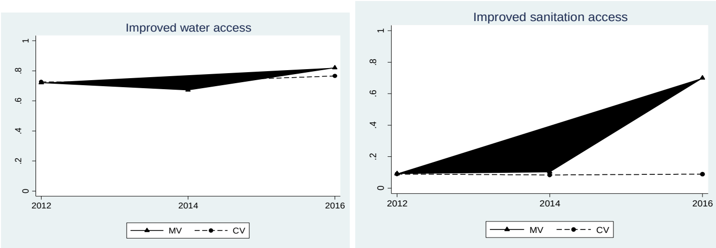
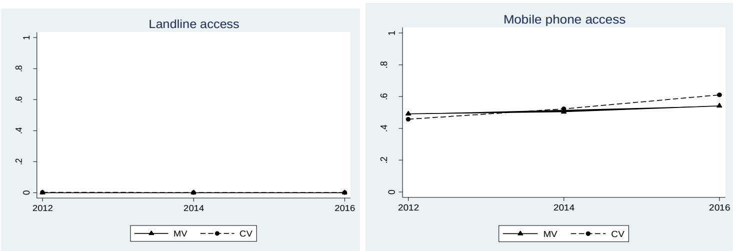


Figure 26 MDG 8: Develop a global partnership for development



The DD impacts of the intervention on the MDGs are shown in Table 22 for each survey year with respect to the baseline and on average (last column) with respect to the baseline. Impacts are expressed as DD: before-and-after changes in the outcomes in the MV areas minus the same changes in the comparator group. They represent the impact of the intervention after removing changes produced by other factors unrelated to the project such as government spending, historical trends or natural disasters. Impacts were calculated from samples matched at the household level using the sub-classification method. When panel data were available, as in the case of the poverty headcount, we estimated a fixed-effects model, while we used cross-sectional estimations in all other cases.

All outcomes are reported as ratios, shares, proportions or rates. Since all MDG indicators are binary outcomes, all the impacts have a simple interpretation as differences in ratios or percentage differences. P-values assessing the statistical significance of the observed effects are reported in parentheses under each DD effect. Statistically significant coefficients are in bold. The first two column of Table 22 show the baseline values of the MDGs and the difference at baseline in the MV area (including a test on the statistical significance of this difference).

Since we are testing multiple hypotheses at the same time, we use critical values for the rejection of the null hypotheses that are more conservative than the usual 5% and 10%. When testing multiple hypotheses there is a good chance of finding effects when there is none. For example, with our 84 different hypotheses and a statistical significance threshold of 5% we would expect at least four 'statistically significant' results when in fact there was no impact. To overcome this problem we assess the significance of the results using the False Discovery Ratio (Efron & Hastie, 2016). In order to do so, we order the P-values in ascending order and index them by $i=1,..,N$ and we reject all the null hypotheses whose P-value is less than $(q/N)*I$, where q is the statistical significance level of choice (in our application we use 0.05 and 0.10 as is common in empirical practice) and N is the total number of hypotheses. We do this separately for the 84 hypotheses relating to year-specific effects and for the 28 hypotheses relating to average effects because effects have to be independent and the 28 average effects are a composite indicator of the 84 yearly effects.

In what follows we discuss the impacts of the intervention on each MDG. The project did not have an impact on any of the indicators of MDG 1 (*eradicating extreme poverty and hunger*). The project did not reduce poverty whether measured by the national poverty line or the national food poverty line. It did however reduce poverty measured using household income data and adjusted by purchasing power parity. The difference of impact on income and expenditure is further discussed in section 8.2. The project did not reduce inequality or improve the distribution of expenditure among the poor as no impact is found on the consumption share of the bottom quintile and on the poverty gap. Finally, there is no impact on the percentage of undernourished children. The project had a positive impact on employment as both the employment to population ratio and the percentage of family workers on own account increased, though none of these effects is statistically significant. Following the reduction in income poverty, there was a reduction in the proportion of the employed that are income-poor.

MDG 2 is *achieving universal primary education*. Primary school attendance was below 70% at baseline and the project increased attendance by 7.7 percentage points in comparison to CV areas, an effect particularly driven by higher attendance in the second and the last year of the intervention. The project did not improve completion rates or adult literacy rates. The latter, it should be noted, was not a goal of the intervention and no specific activity was designed to improve literacy of adults.

MDG 3 is *improving gender equality and empowering women*, which is assessed by gender parity in school and by the share of women in wage employment. The project decreased the parity ratio though the effect is not statistically significant. It should be noted that more girls than boys are attending school in the study area, particularly in the MV villages, as can be seen by the baseline value of the gender parity ratio. The project appears to increase school attendance of boys relatively more than of girls, thus readdressing the existing gender imbalance in favour of girls. We found no impact of the intervention on the percentage of women engaged in wage employment in the non-agricultural sector. Note, however, that a very small fraction of employment is salaried as most individuals are self-employed in agriculture or in non-farm activities.

MDG 4 is *reducing child mortality*, measured by child (under five years) and infant (under one year) mortality and by the rate of measles immunisation. The project does not have a positive impact on any of these indicators. The charts show a clear reduction in mortality rates between baseline and midterm, but further improvements between midterm and endline are small and, more importantly, even larger improvements are occurring in the comparison areas, so that the final DD effect suggests an increase in mortality rates. There was no improvement in the percentage of children immunised against measles. The rate of measles immunisations appears to have decreased in the MV areas against a stable or increasing trend in the control areas. The net outcome is negative in MV areas, though not statistically significant.

Table 22 Impact of MV on the MDGs

MDG	DD Impact 2013	DD Impact 2014	DD Impact 2015	DD impact 2016	DD average impact
Proportion of population below $1.25 (PPP) per day	-1.17 (0.001)	-9.84 (0.013)	-3.95 (0.294)	-9.05 (0.015)	-8.65 (0.002)
Proportion of population below the national poverty line	-0.72 (0.793)	-1.09 (0.745)	0.83 (0.826)	5.567 (0.133)	1.17 (0.676)
Poverty gap ratio	-0.68 (0.804)	-5.89 (0.054)	3.24 (0.218)	1.90 (0.573)	-0.38 (0.869)
Share of poorest quintile in national consumption	1.17 (0.352)	1.49 (0.300)	0.99 (0.392)	-0.01 (0.990)	0.87 (0.321)
Employment to population ratio	2.14 (0.450)	5.22 (0.075)	4.31 (0.083)	0.80 (0.800)	3.06 (0.204)
Proportion of employed people living below $1 (PPP) per day	-13.59 (0.001)	-13.63 (0.001)	-6.60 (0.106)	-10.01 (0.007)	-11.04 (0.000)
Proportion of own account and contributing family workers in total employment	3.89 (0.042)	3.90 (0.046)	4.28 (0.030)	4.04 (0.049)	4.02 (0.037)
Percentage of underweight children under-5		1.03 (0.727)		-2.14 (0.435)	-0.51 (0.821)
Proportion of population below minimum level of dietary energy consumption	-0.84 (0.847)	-9.50 (0.078)	8.81 (0.067)	-0.42 (0.933)	-0.55 (0.885)
Net attendance ratio in primary education	9.56 (0.007)	4.35 (0.252)	3.54 (0.325)	13.48 (0.000)	7.69 (0.015)
Proportion of pupils starting grade 1 who reach last grade of primary	0.90 (0.837)	-1.43 (0.725)	-1.40 (0.741)	-4.12 (0.300)	-1.62 (0.670)
Literacy rate of 15-24 year olds, women and men		-3.36 (0.113)		-0.19 (0.961)	-3.36 (0.313)

Ratio of girls to boys in primary education	-0.29 (0.011)	-0.09 (0.413)	-0.10 (0.420)	-0.26 (0.021)	-0.19 (0.058)
Share of women in wage employment in the non-agricultural sector	-10.97 (0.531)	0.96 (0.960)	-6.92 (0.664)	-14.54 (0.387)	-8.06 (0.545)
Under-5 mortality rate		-20.86 (0.389)			4.12 (0.842)
Infant mortality rate		-8.67 (0.711)			20.22 (0.285)
Proportion of 1-year-old children immunised against measles		-6.45 (0.160)		-3.10 (0.545)	-4.95 (0.182)
Proportion of births attended by skilled health personnel		16.57 (0.001)		39.08 (0.000)	27.00 (0.000)
Contraceptive prevalence rate		5.73 (0.018)		11.48 (0.000)	8.50 (0.000)
Adolescent birth rate					-8.67 (0.269)
Antenatal care coverage		-7.43 (0.129)		2.36 (0.538)	-2.94 (0.468)
Proportion of population aged 15-24 with comprehensive correct knowledge about HIV		0.056 (0.832)		2.41 (0.249)	1.47 (0.474)
Malaria incidence		-4.50 (0.333)		-4.47 (0.345)	-5.53) (0.196)
Proportion of children under-5 sleeping under insecticide treated bed nets		42.88 (0.000)		34.60 (0.000)	39.24 (0.000)
Children under 5 with fever treated with antimalarian		11.13 (0.240)		23.70 (0.023)	15.99 (0.041)
Proportion of the population using an improved drinking water source		-5.89 (0.174)		5.50 (0.129)	-0.27 (0.940)
Proportion of the population using an improved sanitation facility		1.61 (0.444)		61.36 (0.000)	31.04 (0.000)
Fixed telephone subscriptions for 100 inhabitants		0.01 (0.707)		0.01 (0.698)	0.01 (0.675)
Mobile cellular subscriptions for 100 inhabitants		-5.40 (0.374)		-9.96 (0.059)	-7.60 (0.146)

Note: Coefficients are DD estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1% after adjusting for False Discovery Ratio

MDG 5 is *improving maternal health*, which cannot be measured directly in small sample and is approximated by a number of intermediate outcomes, namely the proportion of births attended by a skilled professional, the contraceptive rate, ante-natal coverage, knowledge of HIV and the proportion of children sleeping under a bednet. The project has a large impact on some of these intermediate indicators. The proportion of births attended by professionals and the proportion of children using bednets increased dramatically, while the increase in the proportion of women using contraceptive methods increased substantially. The project did not have an impact on ante-natal visits and on HIV knowledge, however.

MDG 7 is to *ensure environmental sustainability*, assessed by households' access to improved sources of drinking water and use of improved sanitation facilities. The project did not have an impact on sources of drinking water, but had a dramatic impact on access to improved toilet facilities towards the end of project implementation.

MDG 8 is *developing a global partnership for development,* which at the household level is assessed by access to telephone technology. The project did not have an impact on use of landlines, which remained non-existent over the study period in both MV and CV areas. Perhaps a bit surprisingly, given the considerable efforts made by the project in this direction, we found no impact on the use of mobile phones.

Judged against the MDGs, the project can hardly be considered a clear success. In particular, it appears to have failed on achieving a reduction in poverty and hunger. Far from breaking the 'poverty trap' the project does not appear to have reduced poverty and hunger at all. There are, however, some encouraging impacts of the intervention in education and health. The 7.7% effect on primary enrolment is considerable and the dramatic improvement on some intermediate health indicators (births attended by skill professionals, contraception rates and children sleeping under bednets) may be interpreted as predictions of future improvements in maternal and child health that could only be observed in future surveys. Overall, the project appears to have produced mixed results: there is no visible impact on poverty and hunger, a moderate impact on school attendance, a positive impact on intermediate health indicators (attended deliveries, contraceptive use, bednet use), no impact on final health indicators (child mortality), a dramatic impact on access to toilet facilities and no impact on access to water and mobile phone technology.

The impacts on outcome indicators can be mapped back to the project activities designed to affect them. It is perhaps a bit surprising that the project did not reduce poverty considering the resources invested in agriculture and the increase in agricultural income that will be documented in the following sections. Signs of an increase in agricultural activity are visible in Table 22 in the increase in the number of people in work (the coefficients relating to the Proportion of employed people living below $1 (PPP) per day, the Proportion of employed people living below $1 (PPP) per day and the Proportion of own account and contributing family workers in total employment are always positive, though rarely statistically significant). Investments in education appear to have increased primary enrolment, while no activity was designed to improve adult literacy and therefore the absence of any impact is not unexpected. The project promoted girls' school attendance but, in a context where more girls are attending school than boys, it is not surprising that this did not produce the desired

It is also a bit surprising that the project did not have an impact on some outcomes that were explicitly targeted by the intervention such as: child mortality, immunisation rates, ante-natal care, access to drinking water and usage of mobile phones. There is, however, a dramatic impact on the use of improved toilet facilities that emerges towards the end of the intervention. The investments in health services produced the desired effects. More children are sleeping under bednets, more deliveries are attended by skilled professionals and contraceptives are more widely used. It is a bit surprising that ante-natal visits did not increase, though they were already high at the baseline. More worrying is that the project did not have an impact on child mortality. Note that this is not the result of a small sample size being unable to detect a small impact. The data show that child mortality improved more in CV areas than in MV areas during the period considered. Finally, large efforts in the construction of toilet facilities in the last years of the intervention clearly show up in the data but surprisingly the investments in bore-holes did not improve access to water and, despite project efforts, access to mobile communication did not improve either.

The analysis of the impact on the MDGs has a number of limitations, which will be further investigated and analysed in the following section of the report:

- Projects are rarely assessed on impacts on final outcome indicators such as the MDGs. For many of the indicators it could be argued that the size and the timeframe of the intervention was such that it was not able to produce dramatic changes in the short period of time under study. Also, some of the standard MDG indicators like, for example, knowledge of HIV or adult literacy rates, were not explicitly targeted by project activities.
- All MDG outcomes are measured as ratios, shares and percentages, sometimes relying on somewhat arbitrary cut-off points (poverty and undernutrition). This type of analysis precludes observing impacts outside the cut-off points and is blind to distributional effects. For example, the project may have reduced extreme undernutrition or poverty without resulting in an overall reduction in poverty or undernutrition.
- The official MDG indicators are unable to fully detect the impact of the project on living standards. For example, employment-related indicators have little relevance in a place where most individuals do some work at any time. On the other hand, relevant indicators of well-being such as prevalence of anaemia and malaria or cognitive skills and maths test scores are missing.
- There are also a number of issues related to the measurement of some of the indicators. First, only large changes in mortality can observed using small sample sizes. Second, this is an observational study and despite all efforts to build a comparable control group some residual difference may bias the estimation of the project

outcomes. Third, the use of mosquito bednets is seasonally sensitive and may have been mis-measured at the baseline, when project and control surveys were conducted at different times.

- Observed impacts may be dampened by spill-over effects. If benefits extend to control areas, the comparisons are contaminated and the effects underestimated. For example, mothers and children from CV areas may have accessed the health services provided by the intervention thus resulting in large 'project' impacts in CV areas that nullify the comparisons between MV and CV areas.

- Absence of DD impact does not necessarily mean the project did not have an impact. If similar interventions were conducted in other areas, it may simply mean the project is as effective as these other interventions.

Two other fundamental limitations of the analysis conducted above relate to the use of a dashboard approach. First, the approach struggles in making sense of effects on a large number of indicators. The project produced some positive effects but was ineffective in other areas. How can this information be summarised? It is very hard to make a comparison between the MV and CV groups across all indicators at the same time. The obvious solution in this case is the use of a summary index. We will follow this approach in the following section in which we assess the impact of the intervention on a multidimensional index of poverty. Second, the dashboard approach is silent on the mechanisms determining the outcomes. We observed some impacts in some areas and none in other areas but we were unable to explain the determinants of these effects. Section 7 will discuss causal mechanisms operating in the MV project and the following sections will be devoted to analysing outcomes and intermediate outcomes within each sector of intervention in order to shed some light on causal pathways and the constraints faced by the intervention.

5.2. Impact heterogeneity and spill-over effects on the MDGs

In this section we assess the impact of the interventions across two sub-groups (district and gender) and by distance from the MV areas using the stratification of the control group in near and far communities that was imposed by design. Gender refers to sex of the individual for individual-level outcomes (such as mortality or school enrolment) or to the sex of the head of the household for household-level outcomes (such as poverty or access to drinking water). In other words, when the data refer to household-level outcomes such as, for example, poverty, we estimate separately the impact of MV for female-headed households. For district sub-groups we use the original district subdivision that existed at the time of the baseline because the stratification of the sample was based on the two districts of Builsa and West Mamprusi. Spill-over effects are investigated by looking at the impact of the intervention in communities geographically located at a (relatively) short distance from MV communities. The exercise is conducted separately for the districts of Builsa and West Mamprusi because localities in West Mamprusi are more distant from each other and therefore geographic spill-over effects are less likely to occur than in Builsa. The exercise is conducted for all MDG outcomes and the statistical significance of the P-values is calculated, as before, at the thresholds of 5% and 10% using the False Discovery Ratio.

None of the gender differences in impacts is statistically significant. Similarly, none of the differences in effects between the district of Builsa and West Mamprusi is statistically significant. Apparently, the project did not differently affect girls and female-headed households or the populations of two districts. Few of the differences observed in the near communities in comparison to the far-away communities are statistically significant. However, it is difficult to interpret these differences as spill-over effects. The results point to a reduction in poverty in control communities near MV areas, but, since the intervention did not have an impact on poverty in MV areas in the first place, this is more likely the result of other factors rather than spill-overs originating in the project area. When looking at spill-over effects by district we find, as expected, some statistically significant differences in Builsa but not in West Mamprusi. However, some of these differences are again hard to interpret as spill-over effects. Several differences refer again to changes in poverty, which the project was unable to produce in the MV areas in the first place. Other impacts are not in the expected direction, like, for example, a negative impact on the use of bednets or access to drinking water. The only effect that could be possibly interpreted as a positive spill-over effect is an increase in the proportion of birth attended by a skilled professional.

Table 23 Impact of MV on the MDGs: sub-group and spill-over analysis

MDG	DD gender (female)	DD district (Builsa)	DD near	DD near in Builsa	DD near in West Mamprusi
Proportion of population below $1 (PPP) per day	0.073	0.053	-0.065	-0.062	-0.085
	(0.434)	(0.409)	(0.044)	(0.235)	(0.033)
Proportion of population below national poverty line	0.105	0.038	**-0.063***	-0.062	-0.066
	(0.309)	(0.389)	(0.005)	(0.055)	(0.031)
Poverty gap ratio	0.112	-0.027	**-0.066****	**-0.080****	-0.053
	(0.059)	(0.417)	(0.000)	(0.001)	(0.053)
Share of poorest quintile in national consumption	0.033	-0.004	-0.022	**-0.050***	-0.010
	(0.227)	(0.826)	(0.016)	(0.003)	(0.255)
Employment to population ratio	0.030	0.062	0.007	0.012	-0.004
	(0.221)	(0.258)	(0.782)	(0.721)	(0.929)
Proportion of employed people living below $1 (PPP) per day	-0.017	0.160	0.015	0.051	0.011
	(0.419)	(0.026)	(0.813)	(0.492)	(0.862)
Proportion of own account and contributing family workers in total employment	0.026	0.039	0.050	0.059	0.046
	(0.075)	(0.357)	(0.015)	(0.094)	(0.048)
Percentage of underweight children under 5	0.009	-0.077	0.022	-0.022	0.032
	(0.845)	(0.076)	(0.567)	(0.525)	(0.508)
Proportion of population below minimum level of dietary energy consumption	0.169	0.083	**-0.124****	**-0.141***	**-0.122***
	(0.140)	(0.154)	(0.000)	(0.002)	(0.004)
Net enrolment ratio in primary education	-0.054	-0.051	0.054	0.028	0.070
	(0.213)	(0.460)	(0.117)	(0.565)	(0.125)
Proportion of pupils starting Grade 1 who reach last grade of primary	0.025	0.153	0.076	0.103	0.058
	(0.713)	(0.087)	(0.049)	(0.042)	(0.310)
Literacy rate of 15-24 year olds, women and men	0.029	0.070	-0.061	-0.102*	-0.011
	(0.710)	(0.268)	(0.151)	(0.072)	(0.872)
Ratio of girls to boys in primary education	-0.102	-0.162	-0.090	-0.075	-0.102
	(0.830)	(0.409)	(0.278)	(0.560)	(0.364)
Ratio of girls to boys in secondary education	1.528	1.072	-0.439	-2.599	0.826
	(0.754)	(0.617)	(0.722)	(0.544)	(0.222)
Ratio of girls to boys in tertiary education	NA	-2.793	0.241	2.457	0.078
		(0.641)	(0.830)	(0.703)	(0.957)
Share of women in wage employment in non-agricultural sector	0.448	0.310	-0.001	-0.305	-0.035
	(0.337)	(0.142)	(0.995)	(0.342)	(0.799)
Under-5 mortality rate	-0.099	-0.040	0.005	0.021	-0.005
	(0.010)	(0.318)	(0.851)	(0.571)	(0.893)
Infant mortality rate	-0.060	-0.028	0.002	0.004	0.001
	(0.083)	(0.447)	(0.908)	(0.904)	(0.973)
Proportion of 1-year-old children immunised against measles	-0.113	0.111	0.121	0.121	0.123
	(0.116)	(0.150)	(0.052)	(0.152)	(0.134)
Proportion of births attended by skilled health personnel	0.022	-0.067	0.128	**0.275***	-0.009
	(0.775)	(0.536)	(0.033)	(0.002)	(0.855)
Contraceptive prevalence rate	-0.041	0.068*	0.023	0.000	0.030
	(0.645)	(0.056)	(0.220)	(0.993)	(0.117)
Ante-natal care coverage	0.022	0.056	-0.012	-0.009	-0.015
	(0.722)	(0.246)	(0.706)	(0.828)	(0.731)
Proportion of population aged 15-24 with comprehensive correct knowledge about HIV	-0.037	0.021	0.049	0.078	0.027
	(0.402)	(0.565)	(0.044)	(0.122)	(0.165)
Proportion of children under-5 sleeping under insecticide-treated bednets	0.114	-0.198	-0.063	**-0.232***	0.003
	(0.067)	(0.075)	(0.413)	(0.004)	(0.973)
Proportion of population using an improved drinking water source	-0.100	-0.068	-0.055	**-0.109***	-0.005
	(0.141)	(0.169)	(0.038)	(0.007)	(0.896)
Proportion of population using an improved sanitation facility	-0.046	-0.033	-0.022	-0.062	0.022
	(0.391)	(0.335)	(0.271)	(0.019)	(0.450)
Fixed telephone subscriptions for 100 inhabitants	-0.001	0.002	-0.001	-0.003	NA
	(0.544)	(0.608)	(0.634)	(0.620)	
Mobile cellular subscriptions for 100 inhabitants	0.026	-0.118	-0.068	-0.137	-0.017
	(0.596)	(0.170)	(0.153)	(0.065)	(0.767)

Note: Coefficients are DD estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1% after adjusting for False Discovery Ratio

6. Impact of MV on multidimensional poverty

6.1. Impact on multidimensional poverty

Few would question that poverty is the result of deprivations in several dimension and several indices of deprivations have been proposed in the literature. In this section we consider the impact of MV using the Oxford Multidimensional Poverty Index (MPI) of Alkire and Santos (2014). The MPI is a global index of deprivation that was adopted by the UN Development Programme (UNDP) in 2010 for the measurement of global poverty in the yearly Human Development Report series (UNDP, 2010). The dashboard approach employed in the previous section illustrates impacts on a series of indicators, but it is difficult to make sense of the large set of data. An index has the merit of summarising this wealth of information and allowing an easy comparison across groups and over time.

There are a number of other advantages in using the MPI in the evaluation of the MV project. First, the index is theoretically grounded in the capability approach to poverty (Sen, 1992), whereby poverty is the failure to function on a number of welfare dimensions, to some extent independently of opportunities offered by income. The project aims at improving lives along several dimensions, and the use of a metric that captures overall welfare progress appears sensible. Second, income and expenditure measures fail to account for access to public health and education services that may account for much of people's welfare (Bourguignon & Chakravarty, 2003). Not all health and education services are provided through the market (therefore are not captured by income and expenditure figures), and certainly not the services provided by MV. Third, indices such as the MPI have the conceptual advantage of measuring deprivation over several dimensions at the same time. Multidimensional deprivation indices are designed in such a way as to increase when people are failing to meet basic functionings on several dimensions *at the same time*. The index therefore measures the severity of deprivations suffered by individuals in a way that is closer to our common understanding of poverty. Fourth, the index has been adopted by UNDP since 2010 in the Human Development report series, and MPI measurements are available for most countries. This means it is possible to compare multidimensional poverty, and changes, in the study area with poverty and changes in other countries. Finally, from an empirical point of view, the use of an index reduces all welfare dimensions to just one, thereby killing the statistical multiple testing problem.

The MPI is not immune to criticisms. There is some degree of arbitrariness in the selection of the dimensions, the weights and the cut-offs used in its construction, which are not very different from the assumptions made for the construction of more traditional indices, like monetary poverty. It has been observed how the index implicitly makes undesirable trade-offs between welfare dimensions (Martin Ravallion, 2012) and how its construction may provide misleading results in the evaluation of welfare policies (Duclos & Tiberti, 2016). Like the more popular poverty headcount, the MPI score can increase following a policy that redistributes resources from a richer person to a poorer person. In addition, the use of a double cut-offs for the measurement of poverty makes the index sensitive to small changes in the number of deprivations suffered by an individual, which may cause sudden jumps in the index.

The MPI was built to represent deprivation along three fundamental dimensions of welfare that were already employed in the Human Development Index: health, education and living standards. These three dimensions are given equal importance (1/3) and indicators are obtained for each dimension. In particular, a household is deprived in any indicator if (index weight in parenthesis):

- No household member has completed five years of schooling (1/6)
- Any school-age child is not attending school in Years 1 to 8 (1/6)
- Any child has died in the family (1/6)
- Any adult or child for whom there is information is malnourished (1/6)
- The household has no electricity (1/18)

- The household sanitation facility is not improved based on MDG definition (1/18)
- The household does not have access to improved drinking water based on MDG definition (1/18)
- The household has dirt, sand or dung floor (1/18)
- The household cooks with dung, wood or carbon (1/18)
- The household does not own one of the following assets: radio, TV, telephone, bicycle, motorbike, refrigerator, car and truck (1/18)

Our survey questionnaires were modelled to the DHS questionnaire, which in turn were used to build the original MPI. We were therefore able to calculate the index in the same way as the official MPI, with only two exceptions. First, our malnourishment deprivation index is based on child undernutrition only, as our surveys did not measure the nutritional status of mothers (BMI) as is standard in the DHS. Second, we restricted the time over which we calculated child mortality to five years before the survey in order to be able to measure more accurately any changes produced by the intervention. Since not all households have children under five or children of school age, some of the indicators are censored. We follow the same procedure adopted by the MPI (Alkire & Santos, 2014) of considering as not deprived those household for which no information is available to assess their deprivation status. Alternatively, we could remove from the sample the households censored on one dimension but we would end up with a smaller and less representative sample. Our goal here is not to calculate the index with great accuracy but to assess differences between the MV and CV groups, and there is no reason to believe this choice should introduce a bias in the comparison, while it benefits the analysis by granting the use of a larger sample.

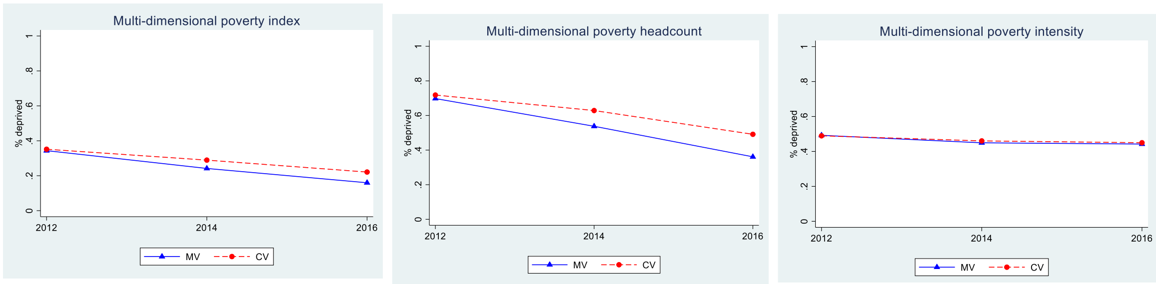
As is standard in the literature, we use the index described above to construct three welfare indicators:

1. The **multidimensional poverty headcount ratio**, also called the **incidence** of poverty. This is the fraction of the population with a deprivation score equal to or larger than one third. Note that deprivation is measured at the household level and that population-level deprivation is obtained by weighting household observations by household size.
2. The average deprivation score of the poor across the population, also called the **adjusted multidimensional poverty index**. This is the MPI, as calculated above, after setting to zero the deprivation score of households that are not multidimensionally poor.
3. The average deprivation score among the poor or the **intensity** or breadth of poverty. This is the average deprivation score across the poor only.

It can be shown that the adjusted MPI is equal to the product of the other two indices: the poverty headcount and the intensity of poverty. In addition, the adjusted MPI has the two desirable properties of sub-group decomposability and dimensional breakdown (Alkire & Santos, 2010), which will be useful in the assessment of the impact of MV. By the sub-group decomposability property, the MPI in a society is equal to the population-weighted average of the MPI in sub-groups of the population – a property that allows an easy comparison of deprivation across sub-groups such as districts or gender of the head of household. By the dimensional breakdown property, the contribution of dimension-specific deprivations to overall MPI deprivation can be calculated, thus providing information on the main sources of deprivation and on where changes are occurring.

The charts in Figure 26 show the three indices in MV and CV areas across the baseline, midterm and endline. The indices cannot be calculated for the in-between years because the adult questionnaire data are not available for those years. We begin our analysis with some observations on the baseline level of multidimensional poverty. There is no difference in multidimensional poverty at baseline between the MV and the CV group. This is obvious from the charts in Figure 26 and it is statistically tested in the second column of Table 24. The poverty headcount in the CV group is 71.8%, while the adjusted MPI is 35.2, which compares with the values of 31.2% and 14.4 for the whole of Ghana reported by Alkire and Santos (2014) for 2008. The study area presents levels of multidimensional poverty that are about twice those observed in the country as a whole.

Figure 27 Multidimensional Poverty Index, incidence and intensity



The charts also show a reduction in the MPI and the poverty headcount in both CV and MV areas, showing that rapid progress in poverty reduction is occurring in the area on several welfare dimensions. Finally, progress appears to occur more rapidly in MV areas than in CV areas. We tested the differences observed in the charts above using regression models and found they are all highly statistically significant (Table 24). The coefficients reported in Table 24 are based on cross-sections rather than panels of households, but estimates using fixed effects and lagged models (not reported) are very similar. The MV produced a considerable reduction in the MPI and on multidimensional poverty. The impact is larger at the endline than at the midterm, pointing to a continuous impact of the intervention across time. Finally, the project decreased multidimensional poverty intensity only marginally, and the impact on the latter was not statistically significant.

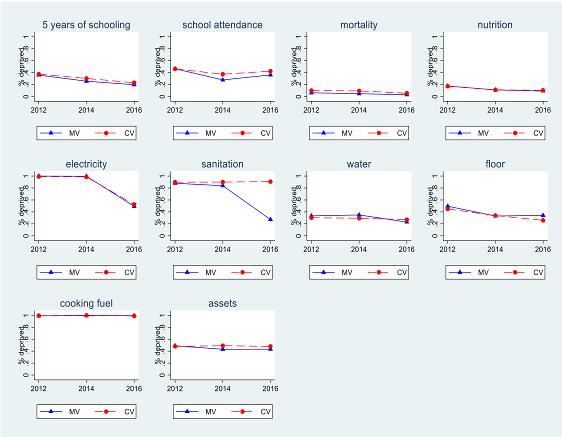
Table 24 Impact of MV on multidimensional poverty

	Baseline in CV areas	Baseline difference in MV	DD impact 2014	DD impact 2016	Average DD impact
Multidimensional poverty index	35.16	-0.81	-3.64*	-5.48**	-4.56**
		(0.793)	(0.070)	(0.017)	(0.019)
Multidimensional poverty incidence	71.82	-2.04	-6.42*	-11.37**	-8.90**
		(0.675)	(0.083)	(0.015)	(0.015)
Multidimensional poverty intensity	48.96	0.24	-1.74	-0.96	-1.50
		(0.857)	(0.116)	(0.470)	(0.115)

Note: Coefficients are DD estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

The MPI is a composite index of deprivations along 10 dimensions and it is useful to see what are the dimensions in which most of the improvement was made. The charts in Figure 27 show the changes on all 10 dimension-specific deprivation indicators. Some impacts are visible in education indicators and sanitation, but no large effects are visible.

Figure 28 Deprivation indices in MV and CV areas



We tested the difference in deprivation indicators at each survey round between MV and CV areas and we found statistically significant differences only in education, child mortality and sanitation. Child mortality, however, was lower in MV area even before the intervention.

Table 25 Deprivation indices in MV and CV areas

	Baseline		Midterm		Endline	
	CV	MV	CV	MV	CV	MV
Years of schooling	37.46	35.86	30.29	25.65	23.12	19.76
		(0.778)		(0.317)		(0.359)
School attendance	46.39	46.50	37.38	27.81**	42.46	36.14
		(0.980)		(0.032)		(0.252)
Child mortality	10.07	6.25**	9.36	4.69**	5.55	3.00**
		(0.021)		(0.004)		(0.017)
Nutrition	17.27	17.54	11.18	11.03	10.38	9.15
		(0.915)		(0.976)		(0.653)
Electricity	99.23	99.98	98.65	99.66	52.47	49.39
		(0.151)		(0.148)		(0.375)
Sanitation	89.78	87.98	89.95	83.86	90.73	27.20***
		(0.638)		(0.214)		(0.000)
Water	30.25	33.34	29.34	34.94	27.10	23.00
		(0.603)		(0.405)		(0.389)
Floor	44.95	49.58	33.95	33.09	25.86	34.16
		(0.563)		(0.839)		(0.171)
Cooking fuel	99.51	99.19	99.94	99.52**	96.97	99.58
		(0.531)		(0.044)		(0.157)
Assets	48.03	49.05	49.16	43.17	47.97	43.31
		(0.825)		(0.200)		(0.187)

Note: Coefficients are DD estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels of the within-survey difference between MV and CV, * is 10%, ** is 5% and *** is 1%

Finally, we look at the contributions of each deprivation indicator to the overall MPI (Table 26). These are obtained by calculating the censored deprivation index for each dimension (therefore setting to zero the deprivation score of households that are not multidimensionally poor), multiplying the index for the weight assigned to the dimension and finally dividing by the size of the adjusted MPI. The percentage contributions are interesting because they tell us what are the main drivers of overall deprivation in the area. Nearly 40% of total deprivation is caused by failure in education, and lack of electricity and sanitation makes another 20%. In other words, concerted investments in electricity, sanitation and education would bring about a large reduction in multidimensional poverty in the area. There are no differences in the contributions of different dimensions between MV and CV areas at the baseline, with the exception of child mortality, which is a less contributing factor in the project area. Interestingly, no change in contribution by any index in MV areas stands out, with the exception of sanitation at the endline. At the end of the project sanitation in MV areas is no longer a large determining factor but contributions by other indices do not change substantially.

Table 26 Contributions of deprivation indices to the MPI

	Baseline		Midterm		Endline	
	CV	MV	CV	MV	CV	MV
Years of schooling	17.45	17.03	17.25	17.53	16.04	17.20
School attendance	21.44	22.22	20.84	18.96	25.73	26.26
Child mortality	4.69	2.99	5.39	2.66	3.47	2.37
Nutrition	7.99	8.06	6.37	7.61	6.71	6.73
Electricity	11.34	11.28	12.04	12.36	7.74	9.94
Sanitation	10.59	10.73	11.31	11.69	11.74	4.72
Water	3.91	4.16	3.94	5.03	4.86	4.32
Floor	5.34	6.47	4.17	5.48	4.14	7.58
Cooking fuel	11.30	11.17	12.06	12.25	12.16	12.56
Assets	5.94	5.90	6.63	6.44	7.41	8.32
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

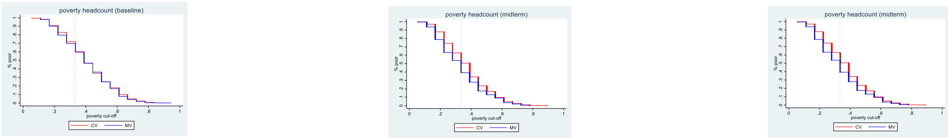
Note: Figures are percent contributions

6.2. Sensitivity analysis

The robustness of results depends on their sensitivity to the parameters used in calculating the indices. We noted that construction of the indices implies an arbitrary choice of a set of dimensions, of weights assigned to the dimensions and of a cut-off point for measuring multidimensional poverty. We accept here the set of dimensions used in the index and focus instead on the sensitivity of the results to the choice of poverty cut-off and the weight given to each dimension.

We assess sensitivity of results to the poverty cut-off using standard dominance analysis. The MPI employs a 1/3 multidimensional poverty cut-off, meaning a household (and all its members) is poor if its deprivation score is equal to or larger than a third. We found that at midterm and endline poverty is lower in MV areas but it would be odd if we found poverty was the same or higher in MV areas if we used another cut-off of, say 1/2 or 1/5. In other words, it would be problematic if the curves representing the percentage of the poor in the two groups at different poverty cut-offs crossed, as this would imply that in MV areas there is more or less poverty than in CV areas depending on the selected poverty cut-off point. The charts in Figure 28 plot the multidimensional poverty headcount for all possible multidimensional poverty lines for the three survey rounds. When the line is 0, everybody is poor. Estimated poverty decreases as we increase the poverty cut-off and when the cut-off is 1 (you must be deprived in all dimensions to be classified as poor) very few households are poor. At baseline the lines largely overlap in MV and CV areas and the poverty headcount is nearly identical for all possible poverty lines. The charts show that multidimensional poverty at midterm and endline is unequivocally lower in MV areas compared with CV areas. At no poverty cut-off the lines are crossing, from which we conclude that the observed impact of MV on multidimensional poverty is independent of the cut-off used to measure poverty incidence.

Figure 29 Multidimensional poverty dominance by survey round

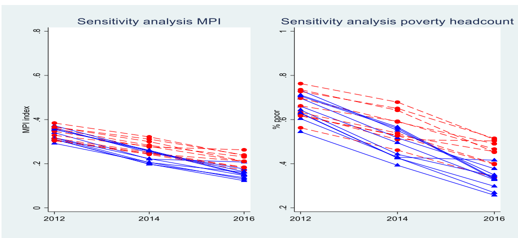


A second robustness concern of the impact analysis is that the results may be driven by changes in one particular welfare indicator, which would point to an impact of MV on a very specific welfare outcome rather than on overall welfare. To test this hypothesis we recalculated the index and poverty incidence after leaving one of the index components out at a time (practically setting to 0 the weight of one component at a time). We did this in a 'nested' way, that is, we reassigned the weight of the missing indicator within its welfare dimension rather than across all dimensions (for example, after excluding nutrition, the weight for mortality becomes 1/3 rather than 1/6, and all other weights remain unchanged).

With 10 dimensions, this exercise produces 10 new poverty estimates. The results are shown in the charts of Figure 29. The blue lines represent poverty in MV areas while the red lines represent poverty in CV areas. Leaving one component out does produce large swings in poverty headcounts in MV and CV areas, but the poverty differences between MV and CV remain unchanged. We tested the difference in the index at each run and we found these to be not smaller than those observed, including all components and always statistically significant a 5%, from which we conclude that the positive impact of MV on multidimensional poverty is not the result of the positive impact on one specific component of the MPI.

We do not extend the sensitivity analysis to more than one component. Since some dimensions have only two components, removing more than one component would lead to calculating the index without the contribution of a dimension like education, health or living standards, which seems to be against the logic of building a multidimensional index in the first place.

Figure 30 Sensitivity of MPI

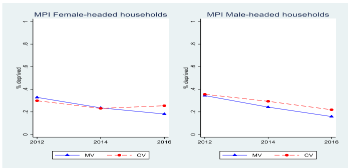


6.3. Impact heterogeneity and spill-over effects on the MPI

In this section we look at whether the impact of MV on the MPI was different for different groups of the population and if there is any sign of spill-over effects. As already mentioned, the MPI has a desirable property of being decomposable across groups. We look at two sub-groups: gender and district of intervention.

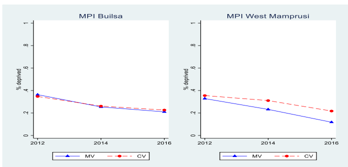
Since the MPI is calculated at the household level, we are not able to calculate an index for male and female individuals. We can, however, calculate indices for male- and female-headed households and compare the differences. The indices for MV and CV areas are in Figure 30 and do seem to show a positive impact in MV areas for female-headed households, though the statistical tests reported in Table 27 show this is not statistically significant. However, much of the impact seems to be the result of a lack of progress among female-headed households in CV areas rather than a faster reduction in poverty among female-headed households in MV areas. Recall that under 10% of households are headed by females and that the samples in the left chart of Figure 30 are small and not truly representative of the population.

Figure 31 MPI of male- and female-headed households



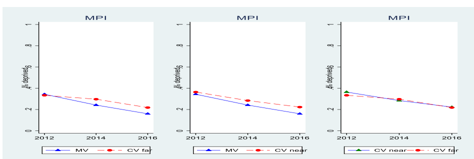
Interestingly, the disaggregation by district seems to suggest the project had a larger impact on the MPI in West Mamprusi than in Builsa, though the differences (Table 27) are not statistically significant.

Figure 32 MPI in Builsa and West Mamprusi



The impact of MV was larger in comparison with far CV areas than in comparison with near CV areas, though this may not be obvious by looking at the charts of Figure 32. The regression results reported in Table 27 show that the decrease in multidimensional poverty was larger in near communities in comparison with far communities and the difference is statistically significant at 10% for the change in the MPI. This could be taken as a possible indication of a spill-over effect.

Figure 33 MPI in CV near and far



In order to shed more light on the presence of spill-over effects, we disaggregate the analysis by district. Recall that near CV communities are much closer to MV areas in Builsa than in West Mamprusi and that we should expect spill-over effects to occur more easily in Builsa than in West Mamprusi. It is perhaps a bit surprising that we find larger changes in near CV areas in comparison with far CV areas for West Mamprusi than for Builsa, as if spill-over effects were larger in West Mamprusi.

Figure 34 MPI in CV near and far areas in Builsa

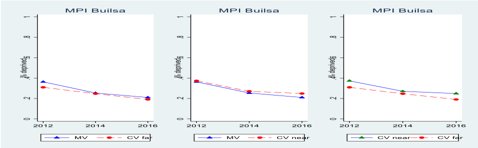


Figure 35 MPI in CV near and far areas in West Mamprusi

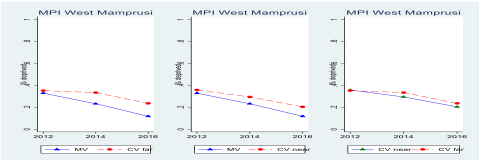


Table 27 Impact of MPI on multidimensional poverty: sub-group and spill-over analysis

	FHH vs. MHH	Builsa vs, West Mamprusi	Near CV vs. far CV	Near CV vs. far CV in Builsa	Near CV vs. far in West Mamprusi
Multidimensional poverty index	-2.18	3.46	-3.39*	-2.24	-4.14*
	(0.698)	(0.416)	(0.076)	(0.463)	(0.091)
Multidimensional poverty incidence	-9.62	7.54	-5.55	2.48	-7.66
	(0.382)	(0.373)	(0.169)	(0.668)	(0.173)
Multidimensional poverty intensity	3.90*	0.68	-1.04	-6.51	-1.01
	(0.099)	(0.706)	(0.425)	(0.710)	(0.544)

Note: Coefficients are DD estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P values in parentheses based on cluster standard errors. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

7. Theory of change and causal mechanisms

So far the analysis has focused on final outcomes, the MDGs and a multidimensional index of aggregate deprivation, rather than on the determinants of the outcomes. As such, the analysis provides limited explanation of why the intervention worked in some areas and not in others. This focus on final outcomes is to a great extent the result of the uniqueness of the MVP model for the following reasons.

First, at the time of the evaluation design the MVP did not have a well-defined TOC of the intervention. One complex TOC by sector of activity developed over several Excel spreadsheets was presented to us before the project start, which included 'pathways' of change for the following sectors: agriculture-hunger, community, education, energy, environment, gender, health and water. However, these pathways consisted of lists of outcomes, outcome indicators and output indicators and 'generic activities' to achieve these outcomes.

For example, the first row of the agriculture-huger pathway included, among others, the following outcomes: *A1.1.1. Increased access to improved seeds, fertilisers in sufficient quantity and quality from private sector agro-dealer networks*, with the associated outcome indicators: *A1.1.1.a Number and % of farming households, disaggregated by gender of HH head, procuring/receiving improved seeds and fertilisers from the private sector; A1.1.1.b Number and % of farmers procuring inputs using vouchers, disaggregated by gender of HH head; A1.1.1.c Quantity and price of improved seeds and fertilisers purchased from the private sector; A1.1.1.d Quantity of improved seeds and fertilisers used on staple foods and other agricultural crops; and A1.1.1.e Number of agro-dealers selling improved seeds and fertilisers to farmers*, output indicators like *A1.1.1.1.a Voucher system developed; A1.1.1.1.b Voucher system includes system for ensuring women's access to agricultural inputs; and A1.1.1.1.c Number of farm demonstration sites established* and a list of generic activities to achieve the outcomes, such as *To develop voucher system, To develop system for ensuring women's access to agricultural inputs* and *To establish farm demonstration sites*. The activities were not clearly related to the outputs and the outcome and there was no explanation of how, or under what conditions, the activities would work, and not even a description of their operation. Even the *Logframe for the Millennium Village in Northern Ghana* (2012) consisted of a series of MDG outcomes and of targets, stating how they would be achieved over the course of the intervention on a percentage basis. The list of activities reported in the pathways appeared to be more a portfolio of potential interventions for the achievement of the indicated outcomes, rather than a selection of activities to implemented in any specific context.

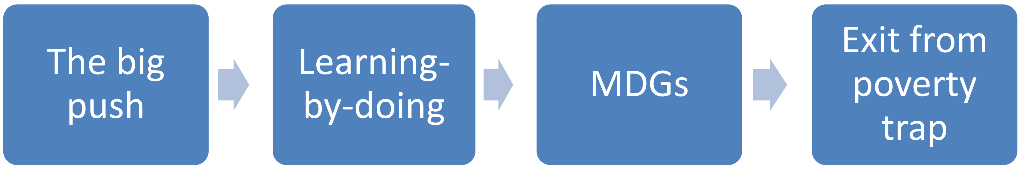
Second, a fully specified TOC for the MV project was bound to be a complicated exercise because the activities promoted by MV are interrelated. In most cases the interrelations are synergistic. For example, higher levels of education could help improving health behaviours, while better adults' health could improve agricultural productivity. But we can also imagine cases in which the activities are producing conflicting outcomes and a sort of 'dis-synergy'. For example, investments in agriculture may require higher labour use by all household members, thus preventing the success of interventions aiming at increasing school attendance. In either case, a full outcome-specific or even sectoral TOC cannot be formulated entirely abstracting from the interventions implemented in other sectors. On the other hand formulating a single TOC for all interventions including all the interrelations would result in a complicated diagram that would ultimately be of little use.

Third, the project aimed at improving a very large number of final outcomes. For each of the 28 MDG indicators discussed above we could formulate a complex TOC, which is the approach impact evaluations normally follow. But formulating and analysing a separate TOC for each outcome proved an impossible task because of the enormous amount effort required. The evaluation team ended up designing sectoral TOCs for agriculture, education, health and infrastructure. But even this disaggregation by sector is extremely complex, and each sector includes several final outcomes that are affected by different activities through a myriad of channels.

Fourth, MV never specified in advance the activities that would be implemented in Northern Ghana to achieve the MDGs. The project approach in this regard was to identify the interventions along the way in a learning-by-doing process. Several activities were tried and discontinued because they were considered unsuccessful while other activities were started from scratch years after the baseline. All of this was unknown at the time of the evaluation design. Designing a questionnaire that could capture impact on intermediate outcomes became guesswork since the activities were not known beforehand. We were also limited in expanding the questionnaire because the collection of data to measure final outcomes alone was already considered too demanding. In short, we were often left without the necessary information to assess the causal mechanisms because appropriate instruments could not be designed.

The Initial Design Document for the evaluation of MVP in Northern Ghana (2012) included an overly simplified TOC of the intervention (Figure 35). Faced with the difficulties in formulating a TOC for the interventions described above, we discussed the TOC of the intervention 'as a package' and focused on the most important, and innovative, aspects of the project.

Figure 36 Schematic theory of change of MVP



The first building block of this TOC was 'the big push'. The MV was designed on the assumption that development is not linear and complex. Sectors are interrelated and development cannot be achieved by acting on one sector at a time. Sustainable development can be achieved only by improving living conditions along several dimensions at the same time, thus exploiting synergies. This is the theory behind 'big push' interventions that have been popular in development economics for a long time. By this approach, development cannot be achieved gradually by means a linear process, but drastically by means of a sudden change in state produced by concerted investments in interrelated sectors. The project would be a package of integrated activities in all sectors of intervention.

As already noted, the MV did not specify in detail the sectors in which investments were to be made. The reason was that the sectors of intervention and the specific activities largely depend on the context of implementation. MV acknowledged the need to design interventions that were fit to the context. MV also promoted local ownership of the project so that the solutions to existing problems could be found by local actors rather than by the project management. Hence, the project adopted an experiential learning approach, whereby the implementers, together with local authorities, would identify the problems and the best way to resolve those problems. It was anticipated that some of these solutions would work while others would not and that the final configuration of project activities would emerge as the result of a learning-by-doing process through testing different activities. It should also be noted that the wide scope of the intervention and the deep involvement of local actors made the MV more a facilitator of change than an actor of change and that much of the project activity was meant to be directed to brokering and leveraging interventions by the government and other donors.

Finally, it was hoped that the learning-by-doing process would help achieve the MDGs, thus showing that a limited integrated investment could do this relatively quickly and at a moderate cost. Breaking the poverty trap, which follows the impact on the MDGs in our diagram (Figure 35), was not an intended goal of the intervention but followed from the theory supporting the intervention. The intervention was not designed to test the poverty trap or any other development theory but breaking the poverty trap and reaching a state of sustainable development is the expected outcome of a 'big push' type intervention addressing the complexity of development. Strictly speaking, however, the goal of MVP was the achievement of the MDGs through an integrated set of interventions.

The acknowledgment of the complexity of development, the concerted investments in multiple sectors, the absence of a TOC or logframe, the testing of activities through a learning-by-doing process, the involvement of local actors at all stages of the intervention and a role for brokers of development rather than implementers are truly innovating aspects of the MVP. Other sections of the report have documented the extent to which the project operated in the intended way, particularly in its relationship with the district authorities and local communities. The qualitative analysis suggests the selection of the interventions in each sector was always based on one of the following strategies:

- Experiential learning. The project would fund feasibility studies (for example in irrigation or market development), which would then be monitored and tested. Successful interventions would be continued and unsuccessful ones discontinued.

- Local solutions. The project would not force the introduction of external off-the-shelf interventions when local solutions were available or when existing solutions could be easily be adapted. For example, health interventions were centred on the deployment of CHWs, which had been developed locally and had been tested over several years at the time MV started activities.

- Quick wins. In some cases, the project relied on interventions that were known to be working in most circumstances, like the distribution of mosquito bednets or fertiliser.

In the next sections we will try to shed some light on the causal mechanisms of the interventions. In doing so we will look at outcomes that are not part of the MDGs but that are closely related to them. In other cases, the questionnaires were sufficiently detailed to collect information on knowledge and attitudes or on inputs, so that some links of impact chains can be uncovered. We will start by looking at the impact of MV on poverty, income and savings and we will then move to impacts within each sector of intervention.

8. Impact on expenditure, income and savings

Did the MV project reduce poverty and did it break the vicious circle of poverty, which is trapping households in the study area? In this section we look more closely at the impact of MV on poverty that was already discussed in the section on the MDGs. We then move to discuss the impact of MV on income and household savings.

8.1. Impact of MV on monetary poverty and expenditure

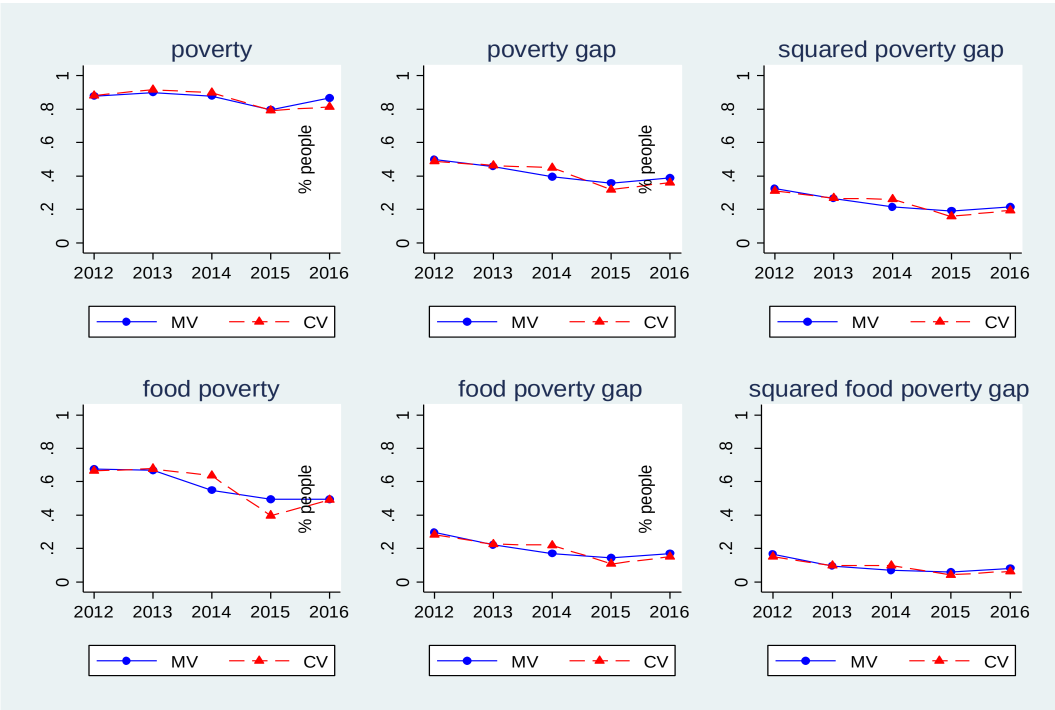
To assess the impact of MV on monetary poverty we use the three most commonly used poverty indices (Martin Ravallion, 2016): the poverty headcount, the poverty gap and the squared poverty gap. The poverty headcount is simply the proportion of individuals in the population whose income is below the national poverty line (a minimum level of expenditure required to allow the consumption of a basic basket of goods defined by the GSS). We consider two poverty lines: a general poverty line, including a basket of basic food and non-food items, and a food poverty line, which only includes the minimum food requirements. The poverty headcount obtained in this way is easy to calculate and to understand but it is not an ideal measure of poverty. Suppose consumption by a group of poor people drops dramatically. The poverty headcount will not change, though welfare of the population has decreased considerably. In other words, the poverty headcount is insensitive to changes in the distribution of poverty among the poor and, as such, is not on its own a good indicator to assess the impact of public policies.

The poverty gap, unlike the poverty headcount, is a poverty measure that is sensitive to the distribution of poverty among the poor. The poverty gap is the percent expenditure gap from the poverty line for the poor averaged across all the population. A reduction in expenditure by the extremely poor will increase the proportionate gap and will increase the poverty gap. The poverty gap therefore increases when the poor get worse off even if the poverty headcount remains unchanged. A drawback of the poverty gap is that it does not capture the severity of poverty. Suppose a monetary transfer is made from a poor household to a less poor households, the poverty gap does not change while the severity of poverty has increased as the poor person became even poorer. One measure that reflects the severity of poverty is the squared poverty gap. The squared poverty gap is calculated in the same way as the poverty gap after squaring all gaps from the poverty line in order to give more weight to larger gaps, those of the extremely poor. One difficulty with the squared poverty gap, and to some extent with the poverty gap, is that they are difficult to interpret. Poverty gap and squared poverty gap rates do not have an obvious meaning unless they are compared across time or across populations. Poverty gap rates become useful, as in our case, when comparing groups or when looking at poverty trends over time. In particular, when making comparison across groups or over time, the poverty headcount tells us about differences and changes in the prevalence of poverty and the poverty gap tells us about differences and changes in the distribution of poverty, while the squared poverty gap says something about differences and changes in the severity of poverty.

Trends of the three poverty indices in the MV and CV areas over time are shown in Figure 36, while the actual rates are reported in Tables 28 and 29 for the general poverty line and for the food poverty line, respectively. These poverty rates are extremely high, being between 80% and 90% in any given year. Nearly all the population is poor and more than 50% of the population is extremely poor for most of the period considered. As a matter of comparison, poverty and food poverty were 24.2% and 7.8%, respectively, in Ghana in 2012/13, and they were equal to 50.4% and 19.3% in the Northern region (Ghana Statistical Service, 2014b). The high rates observed in the study area should not be too surprising, however, considering that our sample is entirely rural (poverty is much higher in rural areas in Ghana) and that the project explicitly selected an extremely poor area for the interventions. Poverty gaps are a function of poverty and therefore are not comparable at different levels of poverty such as the level prevailing in Ghana and in the study area. However, the Upper West region, which reported a poverty rate similar to the one observed in our sample (89.1% in 2005/06), also reported a poverty gap ratio of 50.7%, which is well aligned with the figure observed in our study area. There is no doubt that the population in our study is extremely poor by the standards set by the Ghanaian authorities and that it is to some extent trapped in poverty.

Note that all poverty rates were nearly identical in MV and CV areas at the baseline and that no statistically significant difference was found. Similarly, poverty has remained relatively stable in the study area over the period and has not decreased more rapidly in MV areas compared with CV areas. The MV poverty headcount tracks the CV poverty headcount very closely with the exception of the endline observation. The last round of data collection found an increase in poverty in MV areas compared with CV areas, though this difference is not statistically significant. The poverty gap has decreased in MV areas as well as in CV areas over the four years of the intervention. The distribution of poverty among the poor has improved (meaning it has become less unequal). However, this has not happened more quickly in MV areas than in CV areas with the exception of the observation at the midterm. The squared poverty gap also decreased in both MV and CV areas, pointing to a reduction in the severity of poverty. The reduction in equality among the poor went to the advantage of the poorest. Again, however, there are no large differences between MV and CV areas. The squared poverty gap is lower in MV areas than in CV areas at the midterm, but is larger again at the fourth round. Both poverty gap and squared poverty gaps are nearly identical at the endline.

Figure 37 Poverty indices in MV and CV areas



Unlike overall poverty, food poverty has substantially decreased in the study area over the period considered (Figure 36). However, the reduction was not faster or larger in MV areas in comparison with CV areas. Food poverty was considerably lower in MV areas at the midterm, but again larger in the following round of data collection. The poverty gap and the squared poverty gap follow a similar, and smoother, pattern. Both the distribution and the severity of poverty have somewhat decreased but they have decreased in the same way in MV and CV areas and all the indicators are nearly identical at the endline (Table 28).

Table 28 Poverty indices

	Poverty headcount		Poverty gap		Squared poverty gap	
	MV	CV	MV	CV	MV	CV

Baseline	87.6 (0.836)	88.1	49.7 (0.699)	48.7	32.3 (0.585)	31.0
2nd round	89.9 (0.355)	91.7	45.6 (0.717)	46.3	26.2 (0.773)	26.7
Midterm	87.9 (0.422)	89.8	39.5* (0.066)	44.9	21.4** (0.049)	26.0
4th round	79.6 (0.993)	79.2	35.8 (0.122)	31.8	19.0* (0.061)	15.7
Endline	86.8 (0.125)	81.4	38.8 (0.438)	36.1	21.4 (0.515)	19.4

Note: P values in parentheses based on cluster standard errors. Stars represent statistical significance levels of the within-survey difference between MV and CV, * is 10%, ** is 5% and *** is 1%

Table 29 Food poverty indices

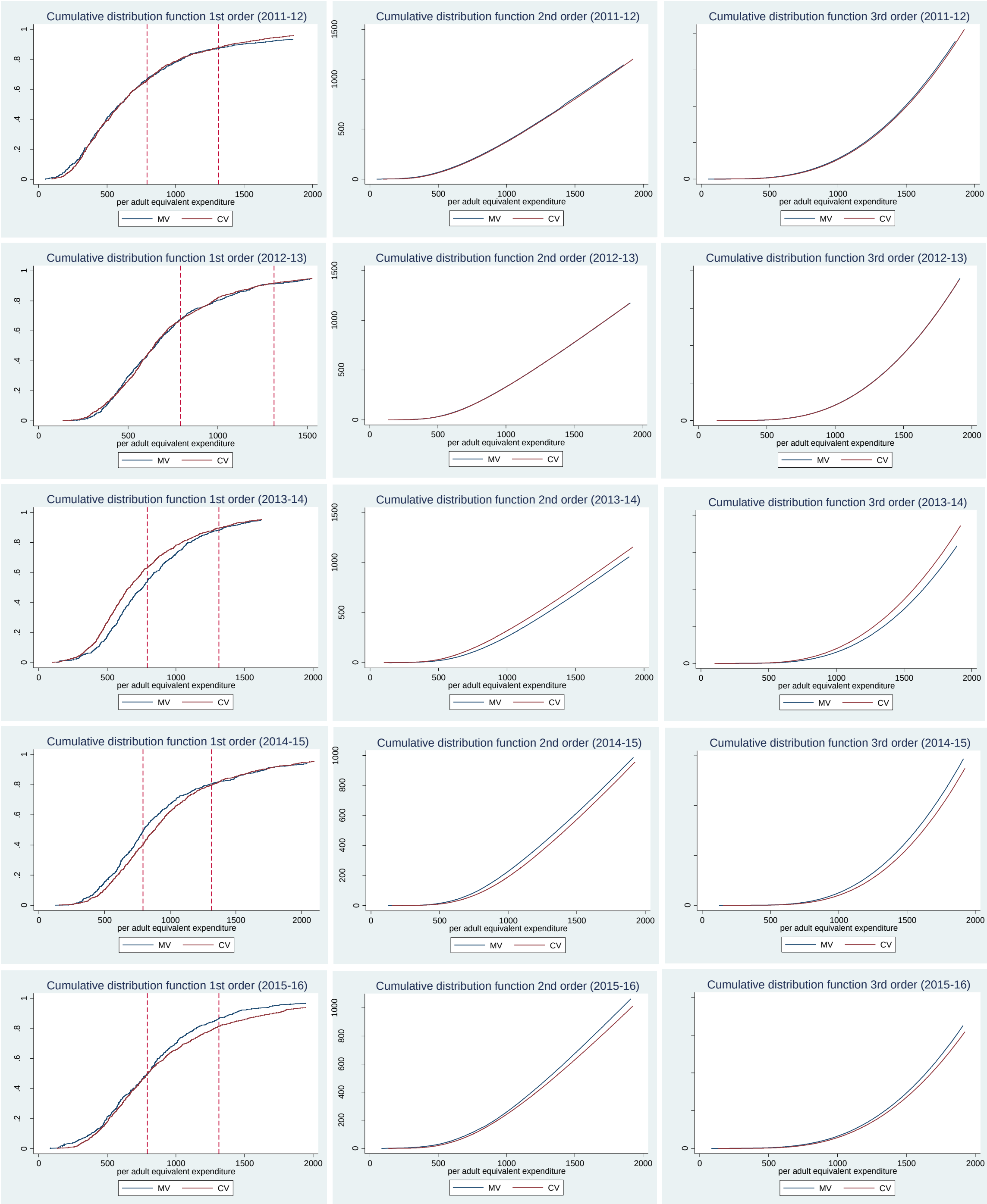
	Poverty headcount		Poverty gap		Squared poverty gap	
	MV	CV	MV	CV	MV	CV
Baseline	67.5 (0.798)	66.5	29.7 (0.572)	28.2	16.5 (0.436)	15.0
2nd round	66.8 (0.790)	67.8	22.1 (0.846)	22.5	9.4 (0.712)	9.8
Midterm	54.7* (0.091)	63.5	16.8** (0.045)	22.0	7.0* (0.064)	9.9
4th round	49.4 (0.044)	39.7	14.3** (0.048)	10.8	5.9* (0.062)	4.2
Endline	49.6 (0.931)	49.1	16.7 (0.633)	15.1	7.9 (0.424)	6.3

Note: P values in parentheses based on cluster standard errors. Stars represent statistical significance levels of the within-survey difference between MV and CV, * is 10%, ** is 5% and *** is 1%

Since poverty headcounts do not reflect changes in the distribution of outcomes, the prevalence of poverty may vary depending on where the poverty line is set. Hence, using two poverty lines, as we have done above (one general and another for food), is always advisable. The concept, however, can be taken further to an infinite number of poverty lines. We can look at poverty rates under all possible poverty lines. This is what is depicted by the cumulative distribution functions of poverty in the left charts of Figure 36. The left charts in the first column of the figure plot (on the y axis) the percentage of poor people as the poverty line increases (on the x axis). The charts in the middle of the figure do the same exercise for the poverty gap, plotting the area under the previous poverty distribution function (that is, the poverty gap times the poverty line), while the charts on the right do the same exercise for the squared poverty gap, plotting the area under the cumulative distribution function of the poverty gap.

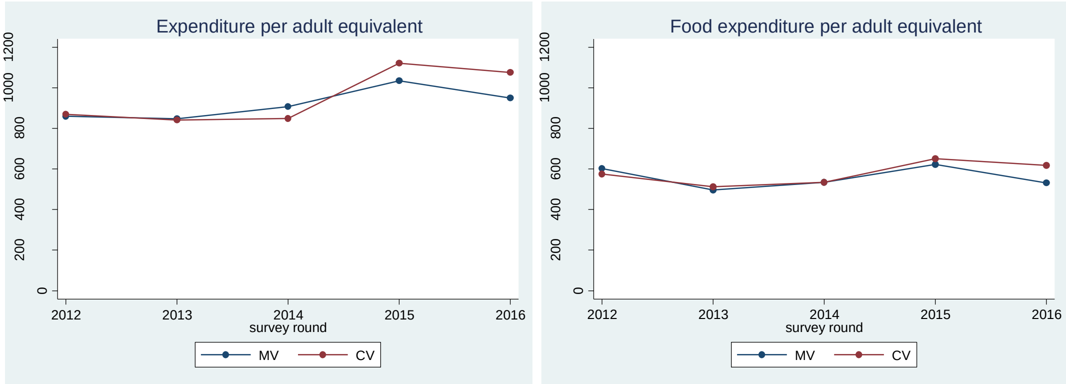
When conducting a dominance analysis of poverty we would like the cumulative distribution functions not to ‘cross’: if they do it means poverty can increase for a given poverty line and decrease for another one. On the other hand, a poverty line analysis is robust when the curves do not cross, meaning that, whatever the poverty line, changes in poverty will consistently occur in the same direction. The poverty distribution functions for all three indices are nearly identical in the first two survey rounds. In the following rounds the curves for MV areas are consistently above or below the curves for CV areas for all indices. The poverty lines are therefore robust and poverty changes are valid regardless of the poverty line set. In other words, there is no poverty line for which it can be found that MV had a major impact on poverty. The last round is a special case as the distribution functions are identical up to the food poverty line and are different thereafter. This leads to the result that for high poverty lines we find differences between MV and CV areas but not for poverty lines below the food poverty line, which is confirming what the data in Tables 28 and 29 told us.

Figure 38 Stochastic dominance analysis of poverty



Since poverty is calculated using consumption data we extend the analysis using the consumption data directly. We run regression models using as dependent variables the log of per adult equivalent consumption and the log of per adult equivalent food consumption and we estimate the project impact using three different model specifications (Tables 30 and 31). Since the dependent variable is in log form, the estimated coefficients can be interpreted as percent differences.

Figure 39 Expenditure and food expenditure in MV and CV areas



The estimated coefficients are in accord with the previous poverty analysis. The intervention did not increase average consumption or food consumption in comparison with the control areas. On the contrary, CV areas appear to have benefited from higher consumption growth, though the effects are rarely significant. The difference occurs in particular during the last two years of the intervention and affects especially food consumption (see also the charts in Figure 38. The coefficients point to a nearly 10% negative impact of MV on consumption in the fourth and fifth years and a negative impact of 15% on food consumption in the final year of the intervention.

Table 30 Impact of MV on per adult equivalent expenditure

	Cross-section	Fixed effects	Lagged model
Average DD effect	-0.017	-0.022	-0.033
	(0.766)	(0.702)	(0.278)
DD effect 2nd year	0.015	0.003	0.001
	(0.835)	(0.970)	(0.987)
DD effect 3rd year	0.090	0.086	0.074
	(0.182)	(0.203)	(0.160)
DD effect 4th year	-0.084	-0.099	-0.101**
	(0.188)	(0.121)	(0.029)
DD effect 5th year	-0.093	-0.084	-0.110**
	(0.225)	(0.279)	(0.043)
Sample size	9,859	9859	7,857

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

Table 31 Impact of MV on per adult equivalent food consumption

	Cross-section	Fixed effects	Lagged model
Average DD effect	-0.045	-0.048	-0.058*
	(0.452)	(0.418)	(0.076)
DD effect 2nd year	-0.021	-0.034	-0.032
	(0.763)	(0.624)	(0.489)
DD effect 3rd year	0.032	0.030	0.019
	(0.680)	(0.698)	(0.793)
DD effect 4th year	-0.059	-0.074	-0.074
	(0.417)	(0.318)	(0.167)
DD effect 5th year	-0.133	-0.118	-0.149**
	(0.146)	(0.188)	(0.027)
Sample size	9,859	9859	7,857

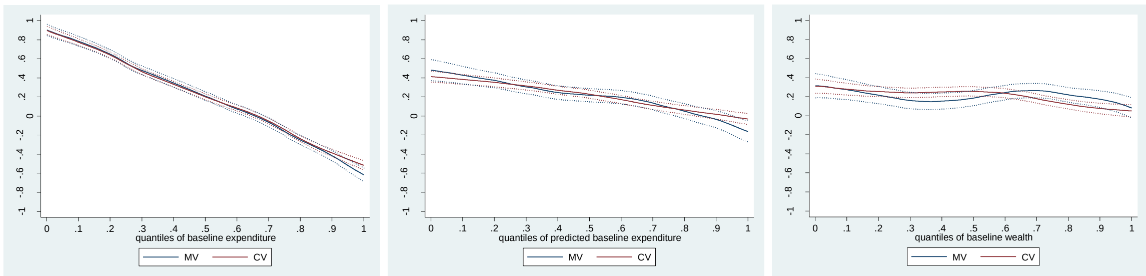
Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

Consumption data are particularly useful to assess changes in the distribution of expenditure. This is often conducted using 'growth incidence curves' (M. Ravallion & Chen, 2003), which plot average growth for increasing percentiles of the consumption distribution. One drawback of growth incidence curves is that they compare different distributions but do not say how growth differs for different groups of people. The latter analysis requires 'individual growth incidence people', whereby average growth in consumption is plotted against the initial consumption, which is done for increasing levels of consumption (Bourguignon, 2011; Fields, Chichello, Freeije, Menendez, & Newhouse, 2003). These curves are better suited to assess whether an intervention has increased expenditure of a specific expenditure group.

One problem with individual growth incidence curves is that they are affected by measurement error. There are two components to this error: classical measurement error and regression to the mean. By classical measurement error, any relationship, positive or negative, between original expenditure levels and future growth is attenuated. By regression to the mean, individuals who suffered a consumption shock at the baseline are correlated to an average growth of the opposite sign in following rounds. The effects of the two types of measurement errors have been worked out by Fields et al. (2003), who show that they produce a growth incidence curve negatively inclined. Notice that a negative slope of the curve means that growth is faster among the poor, which in turn implies that there are no poverty traps as the poor get richer. A negative slope of the individual growth incidence curve is exactly what we observe in our data (see left chart of Figure 39).

This result, however, is affected by the measurement error described above. To remove measurement error, we adopt two different strategies. The first is an instrumental variable approach, whereby we regress baseline expenditure on a number of determinants and we then use the predicted value of expenditure rather than the actual values on the horizontal axis. The second strategy consists of plotting average growth against an indicator of household wealth rather than against expenditure. For the latter exercise, we take the value of household and productive assets combined as an indicator of 'permanent income'. These two strategies are employed in the middle and right charts of Figure 39, respectively. Correcting for measurement error using predicted values of expenditure reduces the slope considerably, while the use of household wealth completely eliminates the slope.

Figure 40 Individual growth incidence curves (expenditure)

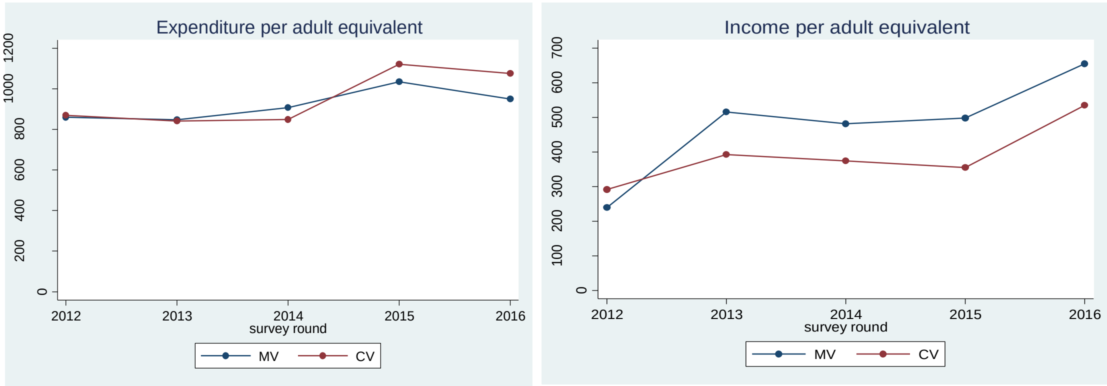


The charts on the right of Figure 39 suggest no poverty trap is in action in the study area. There are no obvious signs that the rich get richer and the poor get poorer. More importantly, there are virtually no differences in the curves for MV and CV. This suggests the MV is not operating to the advantage of the extremely poor and is not breaking a poverty trap in the study area. Based on the lack of impact of the intervention on poverty and on the welfare of the very poor we have to shelve any planned analysis of the impact of the project on the poverty trap as this impact is clearly not occurring.

8.2. Impact of MV on household incomes

The data suggest expenditure did not increase in MV areas after the intervention. Did the same occur to income? Patterns of per adult equivalent consumption and income are displayed alongside in Figure 40. While consumption did not increase in MV areas more than in CV areas, incomes increased substantially more in MV areas. Before commenting on these results, we briefly discuss how consumption and income measures were calculated and how they can be interpreted.

Figure 41 Expenditure and income per adult equivalent



Consumption and income figures were calculated using the methodology adopted by the GSS (Ghana Statistical Service, 2014a). This methodology forms the basis for calculating national official poverty statistics. In the few cases in which following the same procedure was not possible, because of differences in data availability, we adopted best practices available in the literature (Deaton & Zaidi, 2002). The details of these calculations are reported in the Appendix J of the baseline report and we provide a brief summary here. Expenditure is the sum of all goods and services consumed by the household over a year. Goods include food as well as non-food items. Since market development in the area is limited, much of expenditure consists of consumption of own production of food and does not imply monetary transactions. The questionnaire reports the quantities of food items consumed by the household, whether purchased or produced in the farm. The monetary value of production is then imputed using best available prices for each commodity. Non-food items include expenditure in education, health, energy and transport. Purchases of durable goods are not included, but a monetary valuation of their annual value contribution to the household is performed. Similarly, housing rents are not available but, based on household housing conditions, we imputed the monetary value of housing. Our surveys were conducted over a 2-month period and, in order to avoid seasonal bias in reporting of expenditure, most questions (including food expenditures) have a 12-month recall. All expenditures were then adjusted using the regional and monthly consumer price index provided by the GSS. All expenditures are in 2012/13 prices (the year in which the national poverty line was set) and are therefore comparable in real terms across the years.

Income is the sum of all revenues minus costs for each household and each sector of economic activity. As in the case of expenditure, production of goods not sold in the market is valued at best available market prices. The largest component of the income questionnaire collects data on all crops produced in each land plot and the amount of fertiliser, hired labour and other inputs used in the production process. Livestock income includes a valuation of the change in animal stock over 12 months, as well as the value of animal food and non-food goods produced (such as, for example, milk and skins) and the costs incurred in production (such as, for example, fodder and veterinary costs). Each household member reported any income from wage employment for all jobs entertained over the year before the survey. Revenues and costs of all businesses (such as petty trading, small shops, food processing and the like) carried out by household members were reported with reference to a typical month or year. Finally, households reported all monetary transfers received by government programmes, relatives and other private or public donations. As in the case of expenditure, figures were collected using a 12-month recall and were subsequently adjusted for regional and monthly price changes.

Household income and expenditure figures are then divided by the number of 'adult equivalents' in the household rather than by household size. It is believed that per capita figures tend to overstate the extent of poverty and do not allow meaningful comparisons between households of different demographic composition (White & Masset, 2003). This is because consumption needs of children are normally lower than those of adults, at least in deprived areas, and because some household goods can be shared among household members, generating economies of scale in consumption. Methods for adjusting expenditure by the demographic composition of the household vary from simple rules, such as the square root of household size used by the Organisation for Economic Co-operation and Development (OECD), to sophisticated econometric techniques, such as the estimation of Engel's and Rothbart's equivalence scales (Deaton, 1997). Whatever the method used, there is a consensus that, however arbitrary, an adjustment of expenditure and income figures by demographic composition is better than no adjustment at all (Deaton, 1997). In our application, we use the equivalence scale adopted by the GSS, which adjusts people's expenditures in proportion to their nutritional requirements for a given age and sex, in such a way that children weigh less and count only as a fraction of an 'equivalent adult.'

Much of the tradition in the measurement of welfare and poverty has relied on consumption estimates rather than income. However, this has been based more on practical issues of measurement and inertial empirical practice than on a superiority of expenditure over income as a welfare indicator. Economists' opinions are divided between supporters of expenditure – 'There is a strong case for preferring consumption over income in measuring welfare' (Martin Ravallion, 2016) – and supporters of income – 'The multifaceted nature of consumption, and the differing concerns that it evokes, mean that a consumer spending measure is not demonstrably superior to income as an indicator' (Atkinsons, 2015). Both expenditure and income rely on imputations, long (and incomplete) questionnaires and recall and seasonal bias. Some authors believe that practical problems of data collection are more serious for consumption than for income (Deaton, 1997). However, expenditure data are also susceptible to severe errors, and our Benford's analysis of income and expenditure seems to show that approximation, or fabrication, of quantities is more common in reporting expenditure than in reporting income. Theoretical arguments in favour of expenditure over income are also rather weak. An argument is often made of consumption being smoother than income and therefore more representative of 'lifetime' or 'permanent' income, based on the Modigliani-Freedman permanent income hypothesis. The permanent income theory, however, has found little empirical evidence for its validity, and the hypothetical long-term income construct that is the focus of this theory is not necessarily what we are interested in practical and policy applications.

More importantly, income and expenditure indicators measure very different concepts of welfare. Income is concerned with *opportunities*: what can people do with available resources? Meanwhile, expenditure is concerned with the *realisations* of these opportunities: what are actual people's living standards? From our perspective both approaches are equally relevant and welfare comparisons on both indicators are important as long as they are not biased. The charts in Figure 40 show that per adult equivalent expenditure and income were nearly identical in MV and CV areas at baseline. Per adult equivalent expenditure was nearly 10% larger in MV areas at the midterm but then 10% lower in the last two rounds of data collection. The statistical tests reported in Table 30 show that these latter differences were barely statistically significant (note that expenditure is measured in logarithms and therefore coefficients can be interpreted as percent differences). In other words, MV did not have a positive impact on expenditure and there is some evidence that MV areas performed more poorly than CV areas in the last two years of the intervention. Conversely, incomes after the baseline were much larger in MV areas in every single year and the tests in Table 32 show that these differences are consistently statistically significant. Incomes doubled in MV areas in comparison with the baseline, while they increased only by some 50% in the CV areas. The coefficient estimates in Table 32 suggest a difference in difference impact between 40% and 50% depending on the model specification (the dependent variable is measured in units of baseline standard deviations to account for negative values; the coefficients therefore measure changes in standard deviations of income).

Table 32 Impact of MV on per adult equivalent income

	Cross-section	Fixed effects	Lagged model
Average DD effect	0.26**	0.26**	0.205**
	(0.013)	(0.011)	(0.003)
DD effect 2nd year	0.27*	0.27**	0.212*
	(0.036)	(0.036)	(0.039)
DD effect 3rd year	0.22	0.22	0.165
	(0.138)	(0.139)	(0.229)
DD effect 4th year	0.28*	0.29**	0.223*
	(0.051)	(0.044)	(0.039)
DD effect 5th year	0.28**	0.29**	0.222*
	(0.037)	(0.032)	(0.034)
Sample size	9,859	9,859	7,857

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

Since income increased as a result of the project, it is relevant to understand what are the economic activities that generated this improvement and whether they can be traced back to project activities. Income shares for different income-generating activities are reported in Tables 33 and 34 for CV and MV areas separately. At the baseline, more than 80% of income was generated in the agricultural sector. Income generated by micro-enterprises ranged roughly between 10% and 20% depending on the year, while incomes from wage employment and transfers were negligible. There were no significant differences between MV areas at the baseline, and we could not find significant changes in income shares over time either in MV or CV areas, with the exception of a proportional reduction in agricultural income shares and a proportional increase in micro-enterprise shares, which occurred in both MV and CV areas.

Table 33 Income shares by source (CV areas)

Income source	Baseline (2011-12)	Second round (2012-13)	Midterm (2013-14)	Fourth round (2014-15)	Endline (2015-16)
Agricultural	62.9	50.8	45.1	49.9	53.1
Livestock	21.7	26.8	29.7	27.0	25.4
Business	8.2	14.9	17.2	13.7	11.8
Employment	5.2	6.0	6.6	6.7	5.5
Transfers	2.0	1.5	1.5	2.6	4.2
Total	100.0	100.0	100.0	100.0	100.0

Table 34 Income shares by source (MV areas)

Income source	Baseline (2011-12)	Second round (2012-13)	Midterm (2013-14)	Fourth round (2014-15)	Endline (2015-16)
Agricultural	61.6	48.9	44.1	50.4	51.5
Livestock	21.3	26.1	31.3	24.1	24.4
Business	9.6	19.4	17.8	14.5	14.1
Employment	5.6	3.4	5.1	7.6	4.7
Transfers	2.0	2.2	1.7	3.4	5.3
Total	100.0	100.0	100.0	100.0	100.0

Note: In order to calculate the shares we had to set some income values to zero because some households report negative agricultural, livestock and business incomes. We have to assume that negative income shocks are equally distributed in MV and CV area. If they are less frequent in MV areas as a result of the project, then the calculated shares in MV and CV areas are not directly comparable.

Next we test the impact of the intervention on all income sources separately. An important fraction of the sample reports negative incomes resulting from negative income shocks in agriculture. Therefore, incomes cannot be transformed in logarithms to estimate percent changes as we did in the case of consumption. To avoid the estimation of impacts in monetary units, we divide incomes by the standard deviation of income at the baseline. Baseline standard deviations of incomes are reported in Table 35 and can be used to calculate the corresponding impact in cedis. For example, if the standard deviation of income is 600 cedis, an impact of 0.2 standard deviations is equivalent to 120 cedis, which is about 30% of MV baseline income.

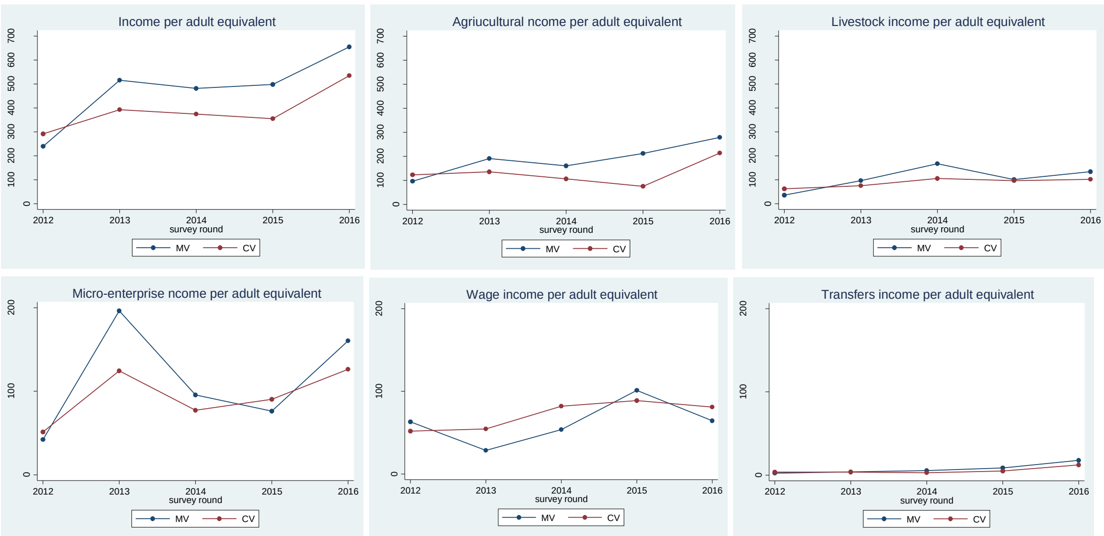
Table 35 Baseline differences in income between MV and CV areas

Income source	MV baseline (2011-12)	CV baseline (2011-12)	Baseline difference
Agricultural	123	96	-26*
	(177)	(211)	(0.099)
Livestock	62	35	-26
	(304)	(354)	(0.241)
Business	51	42	-9
	(236)	(190)	(0.556)
Wages	52	63	11
	(314)	(583)	(0.725)
Transfers	4	2	-2
	(17)	(12)	(0.306)
Total income	291	239	-52
	(592)	(740)	(0.346)

Note: Standard deviations in parentheses. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

Income patterns in MV and CV areas over the project period are illustrated in Figure 41. Incomes from farming are consistently larger in MV areas after the baseline. Livestock incomes are also larger in MV areas, though to a lower extent. Differences in income from micro-enterprises are more erratic, though incomes from micro-enterprises are generally higher in MV areas. Incomes from wage employment are higher in CV areas throughout much of the intervention. Incomes from transfers are consistently large in MV areas, though transfers account for a very small portion of overall income.

Figure 42 Per adult equivalent income by sector in MV and CV areas



Effect sizes and tests of the impact of MV on each income source are reported in Table 36. The table reveals that much of the positive impact on overall income was generated in the agriculture sector and particularly in farming. The impact on livestock income is smaller in size than the impact on farming income. There is also a large impact on income from transfers, though transfers are a negligible component of overall income and therefore contribute very little to the overall impact observed in total income. Impact on incomes from micro-enterprises varies from year to year and is large and statistically significant only at the midterm. Finally, the impact on income from wage employment is negative in all years but the difference is very small and never statistically significant.

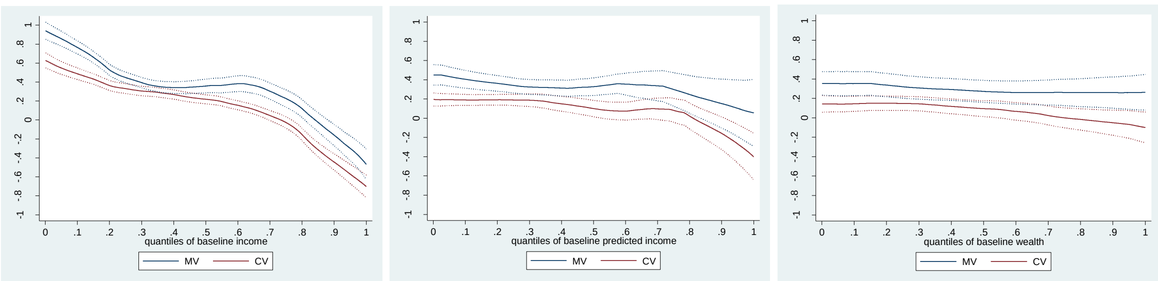
Table 36 Impact of MV on income sources per adult equivalent

	Farming	Livestock	Micro-enterprise	Wages	Transfers
Average DD effect	0.44**	0.08**	0.11	-0.03	0.25***
	(0.001)	(0.049)	(0.313)	(0.519)	(0.000)
DD effect 2nd year	0.34**	0.06	0.32*	-0.06	0.03
	(0.001)	(0.467)	(0.070)	(0.179)	(0.781)
DD effect 3rd year	0.31**	0.18**	0.04	-0.06	0.22**
	(0.013)	(0.004)	(0.893)	(0.248)	(0.027)
DD effect 4th year	0.73**	0.00	-0.10	0.04	0.31**
	(0.001)	(0.966)	(0.505)	(0.477)	(0.008)
DD effect 5th year	0.37*	0.10	0.17	-0.02	0.48*
	(0.068)	(0.195)	(0.422)	(0.703)	(0.057)
Sample size	7,857	7,857	7,857	7,857	7,857

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

A final issue to address is whether the MV is breaking the 'domestic' poverty trap, by which we mean any tendency in the study area to favour accumulation of wealth among the rich. We repeat here the same analysis conducted in the previous section for the estimation of individual growth incidence curves (Figure 42).

Figure 43 Individual growth incidence curves (income)



The unadjusted curves are negatively sloped, pointing to the presence of an income poverty trap. However, after adjusting for measurement error, the individual incidence curves become flat across the distribution of income (or wealth), which implies there is no poverty trap. Incomes grow at the same rate for all households, presumably leaving inequality unchanged. In addition, the distance between the MV and the CV curve appears to be the same across most of the distribution, which implies the project does not improve the income of any specific group in particular, though the difference in the curves is slightly larger at the right of the distribution, suggesting an impact slightly more favourable to the rich than to the poor. All in all, the area does not seem to be affected by a 'domestic' poverty trap and the project does not appear to favour any particular household groups.

8.3. Impact of MV on household savings

It is at first sight puzzling that household incomes are increasing as a result of the intervention, while consumption is not and there is no impact on poverty. In the fourth round report we explained the discrepancy of impact on income and consumption using data providing some support to the operation of the permanent income hypothesis, whereby people save, rather than spend, income changes that are perceived as temporary. We will come back to this issue, but before doing so we need to dispel any doubts about the quality of the income and expenditure data. The analysis of quality of the data using Benford's law suggested that, perhaps surprisingly, income data are collected more accurately than expenditure data. The tabulation of expenditure and incomes by quintiles on a per adult equivalent basis in Table 37, however, is troubling.

Table 37 Income and expenditure quintiles in MVP and GLSS6

	MVP		GLSS6	
	Expenditure	Income	Expenditure	Income
Lowest	256	-152	438	-114
Second	440	58	765	238

Third	642	131	1,135	477
Fourth	930	274	1,731	931
Highest	1,969	1,357	4,232	3,963
Total	846	333	1,810	1,227

Note: Expenditure values are reported in cedis at 2014 prices for the MV baseline data and for a Northern region sample for the GLSS6 data

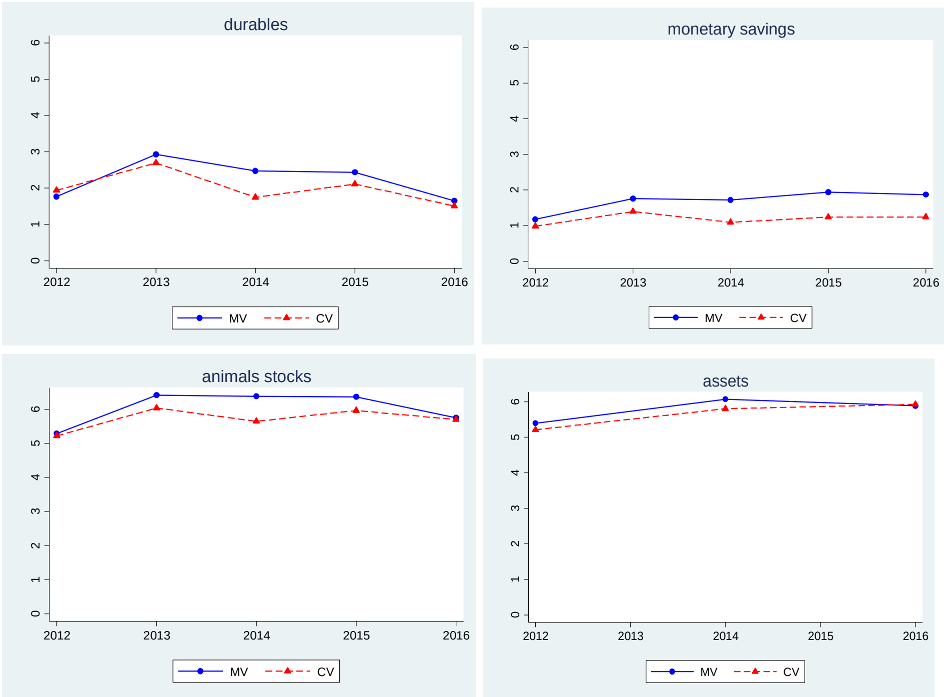
The table shows that reported income is less than 50% of reported expenditure. This is partly an artefact of how income and expenditure figures are constructed. Expenditure is designed to measure welfare and includes items such as the user value of assets owned and the imputed rental value of private homes. However, these items can explain only a small fraction of the difference between income and consumption. A comparison with similar expenditure data for a sample of households from the rural north may shed some light on this. Average expenditure from a similar sample from the GLSS6 (collected in 2012/13, one year after our baseline) turns out to be more than twice the size of expenditure in the study area. It is possible that expenditure in the study area is much smaller than in the GLSS because the area selected by the intervention is particularly deprived. Mean expenditure in the study area is similar to average expenditure in the bottom 40% of expenditure reported by GLSS6 for the rural north. It is also possible that the GLSS6 questionnaire is more detailed than our questionnaire and therefore collecting larger household expenditure. Whatever the reason, it is clear that expenditure is unlikely to be overestimated and it is far more likely that income is underestimated.

Interestingly, the GLSS6 figures also show an average income well below average expenditure, suggesting that underestimation of income is a problem in conducting standard income surveys in Ghana (actually the discrepancy is often found in other survey data in less developed countries). The underestimation, however, appears to be much larger in our survey than in the GLSS6, which may not be a surprise given the difference in the survey instruments used. Looking at income shares may further help understanding the difference between our survey and the GLSS. The GLSS reports an average income share from agriculture of about 40% for the rural north. On the other hand, our data report an income share of agriculture above 80%. This is partly because the study area is more rural than the average of the Northern region of the country, but, possibly, also because the reporting of non-agricultural activities in our data is underestimated. This could partly explain the difference in impacts on expenditure and income. If the MV data report agricultural incomes correctly but underestimate non-agricultural income and if the impact of the intervention occurs mostly on agricultural incomes (as the analysis in the previous section suggests), then the proportional impact on total income produced by MVP is in fact lower than what we are estimating.

A further explanation of the difference between impact on income and consumption was presented in the fourth round evaluation report. Households are investing or saving, rather than spending, their income gains. An important and related question is whether income gains are saved in a precautionary way to protect against future shocks or whether they are invested in productive assets to increase income-generating capacity and therefore future incomes. Data on savings were not directly collected and savings cannot be derived by subtracting consumption from income for the underestimation of income problem described above. We therefore adopt a broad definition of savings and we look at different ways of storing wealth: cash holdings, assets and livestock. Strictly speaking, only cash holdings qualify as savings, as assets and livestock have aspects of consumption in addition to store of wealth or investment, but these are rare in rural Ghana.

All the wealth components considered increased in the MV area after the intervention particularly at the midterm (Figure 43). Purchases of durables, livestock and assets seem to follow an increasing pattern until the midterm and then a decreasing one up to the endline. These charts give some credence to the hypothesis that income gains were spent on durable goods, saved in cash or invested in livestock and assets, at least during the first phase of the project. The charts also show that households did not invest income gains in agricultural assets, except livestock, and that wealth was possibly stored for precautionary reasons.

Figure 44 Durables, monetary savings, animal stocks and assets



We first look at how MV affected expenditures on durable goods and home repairs. Home repairs and purchases of durable goods do not normally enter consumption figures because they involve substantial resources and provide their welfare services over a long period of time. For estimating impacts reported in Table 38, we constructed an expenditure category consisting of home repairs and purchases of the following items: power equipment, computers, mobile phones, furniture, appliances, agricultural equipment and motor and non-motor vehicles, in the 12 months before the survey. We then calculated the value of cash holdings with bank and susu accounts. We calculated the values of all livestock holdings including large livestock, such as cows, donkeys and sheep, and small animals such as guinea fowls, chickens and ducks. Finally, we looked at the value of all asset holdings either for agricultural production (like threshers, hoes and tractors) or for domestic use (like radios, bicycles and furniture). We do not attempt to combine these quantities in a single 'savings' or 'wealth' figure. Durables are measured as a flow while all the other quantities are stocks. Assets were measured only at baseline, midterm and endline. Finally, we believe there is a value at looking at these figures separately.

Table 38 Impact of MV on household assets and savings

	Baseline CV	Baseline diff. MV	Comp. change 2013	Comp. change 2014	Comp. change 2015	Comp. change 2016	Average comp. change
Net borrowing	0.109	0.046	0.222**	0.383**	-0.037	0.097	0.168***
		(0.427)	(0.005)	(0.003)	(0.574)	(0.183)	(0.000)
Bank savings	0.731	-0.005	0.159	0.509***	0.417**	0.280**	0.341**
		(0.972)	(0.288)	(0.000)	(0.003)	(0.027)	(0.001)
Susu savings	0.389	0.223**	0.261**	0.185	0.514**	0.403**	0.340**
		(0.024)	(0.010)	(0.115)	(0.001)	(0.006)	(0.001)
All savings	0.983	0.196	0.341**	0.603***	0.685***	0.574**	0.549***
		(0.271)	(0.045)	(0.000)	(0.000)	(0.001)	(0.000)
Housing goods	1.235	0.020	0.064	0.320**	0.110	0.035	0.132*
		(0.870)	(0.652)	(0.013)	(0.200)	(0.759)	(0.061)
Home appliances	1.199	-0.233**	0.478**	0.799***	0.442***	-0.002	0.431***
		(0.037)	(0.001)	(0.000)	(0.000)	(0.958)	(0.000)
Festival expenses	1.390	-0.181	-0.281**	0.674***	-0.029	0.002	0.090
		(0.217)	(0.032)	(0.000)	(0.814)	(0.990)	(0.407)
All durables	1.940	-0.168	0.231*	0.704***	0.291**	0.112	0.334***
		(0.217)	(0.050)	(0.000)	(0.001)	(0.322)	(0.000)

Household assets	4.856	0.180		0.267**		-0.052	0.109
		(0.242)		(0.004)		(0.685)	(0.266)
Productive assets	2.351	0.152***		0.212**		-0.081	0.070
		(0.000)		(0.011)		(0.508)	(0.460)
All assets							
Small livestock	3.040	0.030	0.311**	0.750***	0.324*	-0.060	0.333**
		(0.877)	(0.040)	(0.000)	(0.085)	(0.725)	(0.007)
Large livestock	4.730	0.027	0.264	0.503**	0.523*	0.465*	0.438*
		(0.920)	(0.341)	(0.035)	(0.090)	(0.095)	(0.051)
All livestock	5.221	0.066	0.366**	0.677***	0.386*	0.022	0.364**
		(0.803)	(0.023)	(0.000)	(0.066)	(0.901)	(0.005)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

The quantities in the figures are in logarithms so that the differences can be interpreted as percent differences. Much of wealth is held in the form of livestock and assets, while monetary savings are of limited size. Unlike assets, livestock is highly liquid as it can be easily traded for cash, particularly small animals, and is an ideal form of saving. The project has a positive impact on net borrowing, meaning households are increasing their stock of debt. This seems in contradiction with a positive change in consumption and stable expenditure (which suggests savings should increase), but the project is promoting loans and the promotion of agricultural activities may encourage farmers to take more loans. There is a large increase in both bank and susu savings accounts (nearly 50% combined), though this type of cash accounts is very rare. There is also a large increase in the purchase of durable goods in the form of both house repairs and various home appliances and goods. The impact on home assets and productive assets is positive but not statistically significant. Finally, MV has a large positive impact on small and large livestock. Note that all these effect sizes are quite large. The effects are also mostly consistent across survey rounds, probably reaching a peak in the year before the midterm survey.

The fact that income increases while expenditure remains unchanged is counterintuitive and deserves some additional explanation. We would expect an extremely deprived population to increase consumption as income increases, but this is not what we observe. From a theoretical perspective this behaviour is consistent with a version of the permanent income hypothesis of consumption. According to this theory, households consume a proportion of their permanent, not current, income and as income increases consumption does not increase, unless the change in income is permanent. Thus, an increase in income can generate two opposite consumption behaviours. An unexpected positive income gain can result in different consumption behaviours depending on how it is perceived. If the increase is perceived as a one-off gain (such as, for example, winning a small lottery amount), the increase is entirely saved and consumption does not change. However, if the increase in income is perceived as permanent (such as, for example, a higher salary following a promotion) consumption does change. The households in our study appear to have perceived the increase in incomes brought about by the intervention as temporary and have therefore not adjusted their consumption upward.

To conclude, a combination of data issues and saving behaviour as predicted by the permanent income hypothesis may explain the apparent contradiction between a sizable impact on income and the absence of an impact on consumption. First, the underestimation of non-agricultural incomes by the MVP survey may have led to an overestimation of the proportionate impact of the intervention on incomes. Second, households may have interpreted the project benefits as transitory and decided to save them in durable goods, bank accounts, home assets and livestock rather than spending them on consumption.

9. Impact of MV on agriculture

MV offered farmers a wide package of interventions. Some of these were piloted and then abandoned when they were found not to be feasible, such as the construction of micro-irrigation structures or marketing development. Other interventions were implemented throughout the project but were only partially successful, such as the formation of agricultural cooperatives or food banks. A number of interventions were implemented more successfully and were widely adopted by farmers, as shown in the section on participation in project activities. These activities include the provision and promotion of agricultural inputs (such as fertiliser, seeds, herbicides and tractor services) and the provision of agricultural training for the promotion of specific crops (maize and beans) and for improving farming practices.

The stated goal of the agricultural interventions is to improve agricultural incomes, food security and the development of an agricultural value chain.[11] Food security is not specifically defined in the project reports and we interpret food security as stable and sufficient availability of food at the household level. Food security was interpreted by the project as a by-product of increased food production and better management of post-harvest losses. Training and cooperatives also were expected to offer partial protection against food scarcity at any particular time. By improving food and nutrition security, the project expected to 'increase incomes… of farming households in the SADA-MVP cluster'.[12] More specifically, MVP aimed to achieve these goals by investing in five key areas in the agriculture sector:[13]

1. Improving delivery of agricultural extension services
2. Improving access to physical agricultural inputs
3. Enhancing agronomic practices
4. Increasing access to agricultural credit
5. Strengthening farmer-based organisations and their linkages to markets

In previous sections we have shown the impact of MV on agricultural incomes. The project was clearly successful in increasing livestock income and farming income. In this section we expand the analysis of the impact of MV on agriculture in three directions. First, we look at the impact of MV on food security. Second, we analyse the impact of MV on land use and cropping patterns. Finally, we decompose the impact on agricultural production through the impact of different inputs such as fertiliser, labour and capital.

9.1. Impact of MV on food security

Food security is a fuzzy concept, which refers to household ability to access food on a stable basis over seasons and years. In the MV survey food security is measured using two questions: whether the household had enough food during the previous 12 months and the number of days the household was without food in the previous month. The percentage of households reporting not having enough food was very high at the baseline (80%) and very similar in MV and CV areas. The intervention had a large impact on household reporting not having enough food in the previous 12 months (Table 39). The fraction of households reporting food insecurity so defined decreased at the midterm by more than 30% with respect to the baseline, and at the endline by 18%. MV also reduced the reported number of days without food in the previous month but the impact is not statistically significant.

Table 39 Impact of MV on food security

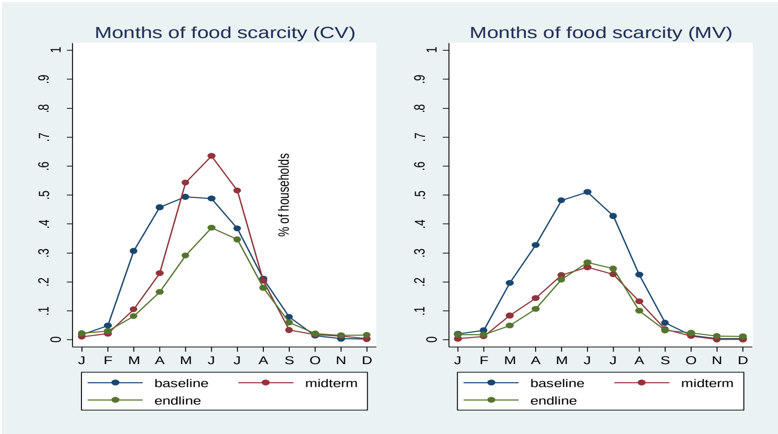
	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Not enough food in past 12 months	82.24	1.36	-32.29***	-18.51***	-25.42***
		(0.663)	(0.000)	(0.000)	(0.000)
Days without food in past month	10.60	-0.79	-1.64	1.08	-0.27
		(0.545)	(0.244)	(0.406)	(0.840)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

[11] 2016, Logframe for the Millennium Villages Accountable Grant Programme, DFID.
[12] 2015, Mid-Year Report on the Millennium Villages Project in Northern Ghana, p. 7.
[13] 2016, Logframe for the Millennium Villages Accountable Grant Programme, DFID.

Households also reported the months of the year in which they were most food insecure. Seasonal stress is highest for food in the months between May and August (Figure 44). Seasonal food insecurity is very similar in MV and CV areas at the baseline. However, MV areas show much lower levels of insecurity at midterm and endline, which is consistent with evidence of an increase in food availability at the household level.

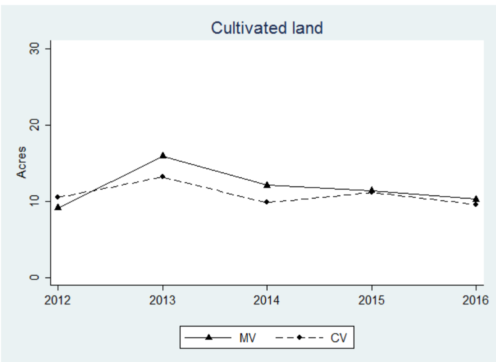
Figure 45 Seasonal food insecurity



9.2. Impact of MV on land use

Land availability is not a major constraint to agricultural production in the area. The average cultivated land at baseline was relatively large (above 10 acres per household) and more land is available for cultivation. Main constraints to agricultural production are the availability of labour, modern inputs (seeds, chemicals and fertiliser) and water. The project encouraged farmers to cultivate larger plots of land. The size of cultivated land increased considerably in MV areas in comparison with CV areas particularly in the second year and at the midterm (Figure 45). In some cases, this occurred by cultivating the existing plots more intensively by means of intercropping. The main crops were maize, beans, rice, millet, sorghum and groundnut.

Figure 46 Cultivated land in MV and CV areas



We disaggregate land use among main crops and find that the project particularly increased the cultivation of maize and beans (Figure 46). The agricultural area devoted to more traditional crops like millet, sorghum and rice did not change while the area cultivated to groundnut decreased. Changes in land cultivated to maize, beans and groundnut are statistically significant (Table 40).

Figure 47 Land cultivated to main crops in MV and CV areas

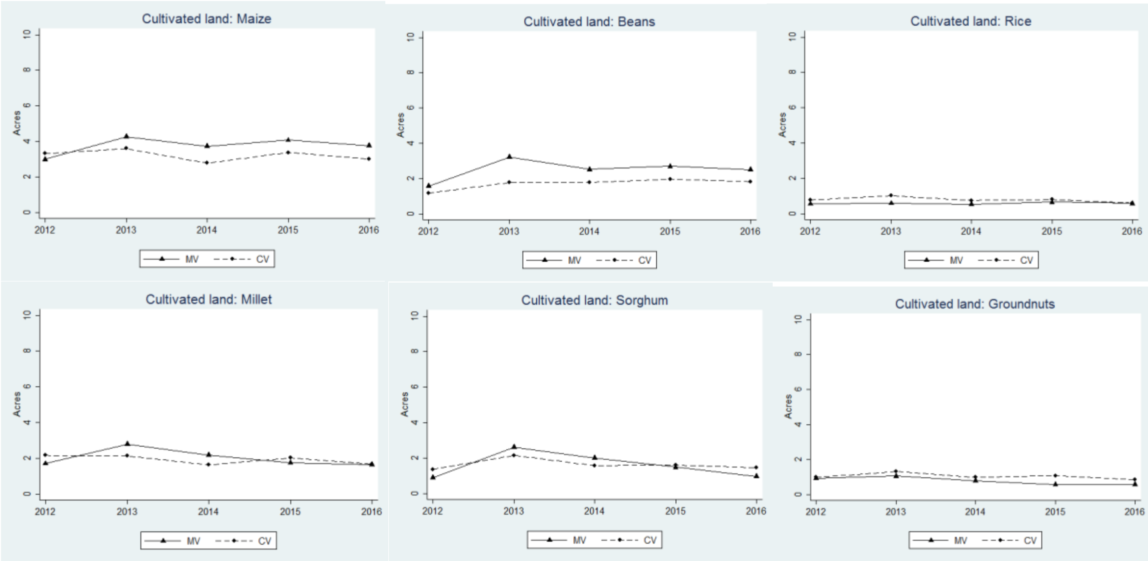


Table 40 Impact of MV on land cultivated to different crops

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2014	DD impact 2015	DD impact 2016	Average DD impact
Cultivated area (acres)	10.50	-1.36**	0.23***	0.21***	0.07	0.07	0.15**
		(0.048)	(0.000)	(0.000)	(0.159)	(0.310)	(0.001)
Maize	3.31	-0.45	0.20**	0.24***	0.22***	0.20***	0.22***
		(0.175)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Beans	1.18	0.40*	0.48***	0.26**	0.27**	0.21*	0.31**
		(0.062)	(0.000)	(0.010)	(0.009)	(0.068)	(0.002)
Millet	2.16	-0.30	0.20*	0.14*	-0.03	0.02	0.08
		(0.453)	(0.068)	(0.094)	(0.722)	(0.878)	(0.369)
Rice	0.70	-0.22	-0.09	-0.04	0.00	-0.01	-0.04
		(0.431)	(0.173)	(0.460)	(0.964)	(0.864)	(0.523)
Sorghum	1.37	-0.47*	0.12	0.08	-0.01	-0.16*	0.01
		(0.068)	(0.304)	(0.329)	(0.875)	(0.053)	(0.927)
Groundnut	0.98	-0.06	-0.05	-0.07	-0.18***	-0.10***	-010**
		(0.563)	(0.300)	(0.128)	(0.000)	(0.009)	(0.008)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

9.3. Impact of MV on agricultural productivity

To tackle the agricultural problems faced in northern Ghana, MVP employed a generic set of activities aimed at achieving 'quick wins' by delivering inputs, subsidising improved seeds of high-yielding crop varieties or hybrids, training farmers on agronomic practices to eliminate 'hunger months', forming cooperatives and developing food storage options and markets. The interventions were not explicitly connected to each other in a causal chain form, though they were supposed to contribute to improving profits in various ways. Agricultural profits are expressed as output prices (p), times output quantities produced (q), minus input prices (w), times input quantities used (x):

$$\pi = pq - wx$$

The project aimed at increasing farm profits in several ways. First, it provided farmers with agricultural inputs such as seeds, fertiliser, water, tractor services and land preparation at a below market rate and, in some cases, free of charge. Seeds and fertiliser were either donated or provided through loans made on concessional terms or with very low repayment rates (i.e. heavily subsidised). Therefore, they affect the quantity of output produced directly (q), while mildly affecting the quantities purchased (x). The provision of other inputs that are complementary to seed and fertiliser, like water and tractor services, was less systematic and less successful.

Farmers are allowed to rent small tractors at below market rates, but only 10 tractors are available to loan and the subsidy amounts to 20% of the market cost. The project helps farmers prepare land for cultivation, but only for a limited number of rice plots. Finally, studies were conducted for micro-irrigation projects but never implemented because they were made redundant by government plans to build a new dam on the White Volta River, which will positively affect the whole area under MVP. Additionally, to ensure agricultural extension agents (AEAs) can access the rural communities they serve, MVP provides motorbikes and fuel stipends.[14]

Second, the project promoted agricultural extension services. The MVP helped hire new AEAs, adding to those already employed by the government. In addition, AEAs were given training and basic tools. They were supervised to increase efficiency and time actually spent in the communities. AEAs worked through more than 150 'lead farmers', who were selected in each community based on skills and motivation and were in charge of managing farmer groups of 15-20 members. Lead farmers were equipped with tools and training and charged with the task of training their farmers' group. Training relied heavily on farm visits and demonstration plots and included sessions on planting, land preparation, weed control, harvesting, integrated soil fertility management and post-harvest management. The MVP expects training to increase profits by increasing farm productivity through an increase in agricultural production (*q*).

Third, the project promoted the formation of cooperatives and market development. MVP conducted a number of studies on agricultural systems and value chains to inform the selection of promising new crops and to improve market access. Large buyers for farmers' produce were identified and farmers received training on market quality standards and requirements. Mango, maize, millet and acacia were identified as promising new crops and farmers were given saplings and training to grow them. Market development initiatives were expected to improve profits by giving farmers access to better prices and promoting the production of higher-value crops (*p*). As a vehicle to achieving this objective, the project has given great attention to organising farmers through the formation and capacity-building of cooperatives. Farmers' cooperatives were formed in each community at the onset of the project and two cooperative officers were hired to support them. Cooperatives were formed following Ghanaian legislation on cooperatives. Cooperative members were trained by the project and the cooperative structure is used to channel agricultural loans to farmers. Many benefits may be generated by cooperatives, including the opportunity for farmers to spread agricultural risk among members, increased access to credit via collective responsibility of loans and increased negotiating power with traders in determining input and output prices. Cooperatives can help increase profits in several ways.

Fourth, the project invested in minimising post-harvest losses. Some of the losses are resolved by training farmers on proper harvest times. However, other losses are the result of improper storage methods or the absence of storage facilities. The project therefore rehabilitated warehouses or built entirely new storage facilities. Improved storage has an immediate impact on quantities of output sold, as losses are reduced, but also helps prices as farmers have the opportunity to sell their produce when prices are more favourable. This intervention therefore helps profits by positively affecting output quantities and prices.

Finally, it should be noted that MVP consisted of a varied package of interventions in multiple sectors including education, health and infrastructure. Investments in other areas may benefit agricultural profits indirectly. For example, improvements in the health of the population may result in a reduction of working time lost to sickness and a general increase in workers' productivity. Similarly, the rehabilitation of roads or construction of new roads can improve access to markets and help change prices to farmers' advantage. Other interventions may improve all components of profits by increasing total farm productivity (*q*), changing prices favourably for farmers (*p* and *w*) and providing better access to inputs (*x*).

To assess the impact of MVP on agricultural production we first investigate the impact on the quantity produced (*q*) rather than on profits for two main reasons: 1) most of the impact on profits should occur through changes in the quantities produced. The input package increases input use, while training and other interventions increase the productivity of inputs. Impacts on profits occurring through prices are less likely because they mostly rely on other interventions and on market development initiatives, which were not particularly successful; and 2) the

[14] 2014, Annual Report on the Millennium Villages Project in Northern Ghana.

estimation of a profit function requires good data on input prices, and our data on wages and other inputs are incomplete and inaccurate. Local community-level prices were not collected through market surveys but by interviewing 'knowledgeable' individuals in the communities.

The simplest specification for the agricultural production function is the Cobb-Douglas form:

$$q = Ax^{\alpha}z^{\beta}$$

where q is the quantity of output produced, x's are variable inputs such as fertiliser, seeds and labour and z's are fixed inputs such as land and productive capital. The α and β parameters measure production elasticities, that is the percent increase in quantity produced for a percent change in the quantity of input used. Finally, A is total factor productivity or 'disembodied' technical efficiency, that is any contributions to production that are not embodied in the inputs included in x and z.

One advantage of the Cobb-Douglas form is that it can be easily estimated with OLS using a logarithmic transformation:

$$\ln q_i = \ln A + \sum_{j}^{n} \alpha_j \ln x_{ji} + \sum_{m}^{k} \beta_m \ln z_{mi} + \varepsilon$$

This specification allows for the separate estimate of the contributions to production of variable and fixed inputs (parameters αand β), and the contribution of any other factor (parameter A) or total factor productivity.

However, we cannot estimate agricultural production as the sum of quantities produced of each crop. Our farmers produce a variety of agricultural goods that cannot be simply added up unless they are transformed in values by multiplying quantities produced by their prices. The dependent variable is therefore the value of the agricultural production, that is quantities of each crop produced multiplied by its price ($v=qp$). Production values are not quantities and make the dependent variable sensitive to price variations, as a higher production value may simply reflect the production of a higher-value crop. To remove price effects we normalise the output value by a price index calculated at the household level. The price index is a geometric mean of median village-level crop prices weighted by the farm-level production (value) share of each crop. The dependent variable of our production function is a normalised production value. We divide production values by the price index P, which is calculated in the following way using 22 village-level crop prices:

$$P_i = \prod_{j}^{n} P_j^{\frac{v_{ji}}{\sum v_{ji}}}$$

Variable inputs (x) are quantities of inputs used over the previous 12-month agricultural year. They include kilograms of seeds, chemical fertiliser, herbicides and pesticides used in all cultivated plots and the number of days of own labour and hired labour in agriculture over the previous 12 months. Fixed inputs include land and capital. Land is measured in hectares of cultivated land while capital assets are measured as the value of the sum of the following production assets: animals (oxen, horses and donkeys), animal-drawn cart, tractor, plough, hoe, axel, shovel, spraying machine, sickle and power tiller.

We now show how an application of the standard Oaxaca decomposition to a DD analysis of the production function can help us understand how MVP affects agricultural production. After this exercise it will be possible to separate the observed project impact on agricultural output in 1) a component resulting from changes in input used; 2) a component resulting from changes in returns to inputs (productivities); and 3) an otherwise unexplained component. It should be recalled that, based on our understanding of the way the project operates that was described in the previous section, changes in inputs are mostly determined by the delivery of the package of inputs; changes in productivities are mainly obtained through farmer training; and changes in total factor productivity are determined by changes in other MV interventions. In order to explain impacts resulting from changes in input use by the delivery of the input package, impacts determined by changes in returns to input can be attributed to farmer training; any other unexplained impact can be attributed to an overall impact of the project on productivity via improvements in health or access to markets.

To illustrate our application of the Oaxaca decomposition to DD analysis, we start by taking first differences of the production function over two periods, thus relating changes in agricultural output to changes in inputs. We do this separately for the project (1) and control observations (0). In order to simplify notation, we consider only a change over two periods, we ignore logarithms and we consider a single input (*x*). We employ the difference operator *d* to express changes in variables from one period to the next. The 'differenced' production functions in MV and CV areas, respectively, are:

$$dq_1 = A_1 + \beta_1 dx_1$$

$$dq_0 = A_0 + \beta_0 dx_0$$

Subtracting the first expression from the second gives the DD estimator of programme impact:

$$dq_1 - dq_0 = A_1 + \beta_1 dx_1 - A_0 + \beta_0 dx_0$$

By adding and subtracting $\beta_0 dx_1$ we obtain:

$$dq_1 - dq_0 = A_1 + \beta_1 dx_1 - A_0 + \beta_0 dx_0 + \beta_0 dx_1 - \beta_0 dx_1$$

which simplifies to the familiar Oaxaca decomposition:

$$dq_1 - dq_0 = (A_1 - A_0) + \beta_0(dx_1 - dx_0) + (\beta_1 - \beta_0)dx_1$$

which decomposes the DD effect into 1) a component brought about by the difference in input changes $\beta_0(dx_1 - dx_0)$; 2) a component brought about by the difference in input productivities $(\beta_1 - \beta_0)dx_1$; and 3) a component otherwise unexplained resulting from changes determined by the project $(A_1 - A_0)$. Estimation over multiple periods simply requires the inclusion of time variables for each survey round, which capture otherwise unexplained changes from one round to the other. In our analysis we use four survey rounds and three year-to-year changes and therefore include two time dummy variables in the estimated regressions.

We start by estimating the impact of the MVP intervention on agricultural output. The outcome considered is the value of agricultural production normalised by a price index. The project has a large impact on agricultural output. The average effect is 0.38, meaning that, on average, every year agricultural output is 38% higher than at baseline in MV areas in comparison with CV areas. The disaggregation of the impact by year shows that the impact of the intervention was particularly strong in the third year and similarly good in the second and the fourth year.

We then estimate the baseline production function described above separately for MV and CV areas (Table 41). We test the difference in the coefficients of the production functions of MV and CV and find that the production functions in the two areas are structurally equivalent at the baseline. None of the T-statistics testing the equality of the coefficients across the two equations are significantly different at the 5% level (third column of Table 41) and a F-test of joint equality of all coefficients is not rejected.

Table 41 Agricultural production function at the baseline

	MV areas	CV areas	Test of difference
Fertiliser	0.020	0.036***	1.12
	(0.115)	(0.000)	(0.264)
Seeds	0.092**	0.166***	1.70
	(0.003)	(0.000)	(0.092)
Herbicides	0.030**	0.031**	0.06
	(0.001)	(0.013)	(0.950)
Pesticides	0.035**	0.041***	0.31
	(0.039)	(0.000)	(0.756)
Labour	0.314**	0.308***	0.06
	(0.003)	(0.000)	(0.955)
Land	0.104**	0.099**	0.07
	(0.042)	(0.006)	(0.942)

Capital	0.017	0.009	0.47
	(0.259)	(0.281)	(0.639)
Constant	4.698	4.664	0.05
	(0.259)	(0.281)	(0.958)
R-square	0.415	0.466	
Observations	665	1,242	

Note: Coefficients of OLS regression adjusted by IPW method. P-values in parentheses based on cluster-adjusted standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance. F-test of joint significance of all joint coefficients was rejected (F=0.98, P-value=0.453)

Finally, we decompose the impact of the intervention on agricultural output using the standard Oaxaca decomposition approach described above. The results are shown in Table 42. The first column shows the impact of the intervention on agricultural output without including any control variables and estimates an average impact of the project using a first difference estimator. We find an average impact of 0.38 of the intervention. Since the dependent variable is in logarithms, this is roughly equivalent to a 38% increase in agricultural productivity. The second and the third column of Table 42 estimate the production function for the MV areas and the CV areas separately. The results of these regressions are the basis for performing the decomposition of effects, which are presented in the fourth and fifth columns. The fifth column shows the changes in agricultural output that are explained by changes in input use. These effects represent how much the output in CV areas would increase if in CV areas input use had to increase in the same way as in MV areas. The majority of input coefficients are statistically significant and the input increases together explain 74% of the change in productivity. Fertiliser, seeds, land, tractor rents and other rents (animal and machinery) appear to make the largest contributions to production change in MV areas. The fourth column shows the changes in output that are not explained by changes in inputs. Unexplained changes can be subdivided in changes in returns to inputs and other unexplained changes (MV and time dummy variables in the fourth column). Changes in returns to factors are very small, sometimes have a negative sign (pointing to a decrease in input productivity in MV areas) and are never statistically significant. Other unexplained changes occurring over time have a relatively small impact on the increase in agricultural output and the coefficients are not statistically significant.

Table 42 Decomposition of impact of MV on agricultural production

	DD effect without control variables	Production function in MV areas	Production function in CV areas	Changes in factor productivities	Changes in factors
MV total difference	0.381***				
	(0.000)				
T2	-0.196***	-0.097	-0.246***	0.037	
	(0.000)	(0.170)	(0.000)	(0.199)	
T3	-0.174***	-0.107	-0.050	-0.014	
	(0.000)	(0.131)	(0.312)	(0.541)	
T4	-0.249***	0.049	-0.119**	0.042	
	(0.000)	(0.493)	(0.015)	(0.152)	
Fertiliser		0.004	0.011***	-0.012	0.028**
		(0.241)	(0.000)	(0.503)	(0.007)
Seeds		0.096***	0.159***	0.000	0.078**
		(0.000)	(0.000)	(0.999)	(0.033)
Herbicides		0.007	-0.003	0.029	-0.004
		(0.228)	(0.368)	(0.544)	(0.579)
Pesticides		0.033***	0.048***	-0.019	0.021
		(0.000)	(0.000)	(0.508)	(0.245)
Labour		0.284***	0.221***	0.010	0.025
		(0.000)	(0.000)	(0.572)	(0.141)
Land		0.106***	0.134***	-0.024	0.083**
		(0.000)	(0.000)	(0.666)	(0.037)
Capital		0.023**	0.014***	0.019	0.007
		(0.003)	(0.006)	(0.644)	(0.203)
Tractor		0.042***	0.030***	0.013	0.013*
		(0.000)	(0.000)	(0.650)	(0.056)
Other rents		0.023**	0.030***	-0.008	0.031**

		(0.029)	(0.000)	(0.784)	(0.006)
Component of difference				0.098 (0.237)	0.283*** (0.000)
R-square	0,018	0.359	0.369		
Observations	7,652	2,644	4,988		

Note: Coefficients of OLS regression adjusted by IPW method. P-values in parentheses based on cluster-adjusted standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

This analysis suggests that much, if not all, the observed improvement in agricultural output observed in MV areas is the result of an increase in input use, in particular of fertiliser, seeds, land, tractor and animal services. We do not observe an improvement in the productivity of inputs, which would point to an impact of agricultural training beyond what is incorporated in the use of inputs and the cultivation of new crops. None of the inputs productivities in the MV areas increases in comparison with in CV areas. Finally, there is a residual unexplained positive effect of the project on agricultural output that is not embodied in changes in inputs. These are changes attributable to other MV interventions, such as roads or irrigation. The effect, however, is not statistically significant and not very large. The fact that the increase in agricultural output is mainly the result of an increase in the use of inputs provided by the project raises some important questions about the sustainability of the intervention. It is not certain that high levels of input use will be maintained by farmers once project support is discontinued.

10. Impact of MV on children's health

10.1. Impact on child mortality

Mortality rates have been decreasing in Ghana over the past 30 years. The reduction in mortality rates has been much faster in the north than in the south of the country. In 1985 in Northern Ghana one in four children would die before their fifth birthday. In 2011 the same probability was one in ten, not too different from the probability in Southern Ghana (see Figure 47).

Figure 48 Under-five mortality rates in Northern and Southern Ghana 1985-2011

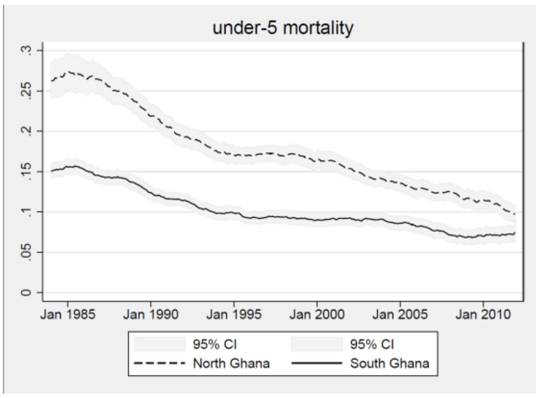


Figure 47 was built by pooling data from six different DHS: 1988, 1993, 1998, 2003, 2008 and 2011. Mortality rates were calculated for each of 300 months before the survey in 2011 exploiting mothers' retrospective recall of births and deaths. When pooling data from different weighted samples a problem arises about the use of the existing sampling weights. We used here the simplest reweighting scheme consisting of adjusting the sampling weight in each survey by the sample size contribution of the survey to the sample of pooled surveys. For example, the sampling weights of the 1988 survey with sample size N_{88} are obtained by multiplying the sampling weights of the 1988 dataset by the ratio N_{88}/N, where N is the sample size of all pooled datasets. Mortality rates were calculated using the synthetic cohort probability method used by the DHS. The rates are calculated by means of a Stata package developed by the authors (Masset, 2016), which reproduces the DHS rates exactly and allows testing differences between groups and plotting trends.

Child mortality is the ultimate health indicator, and several activities had the goal of increasing children survival rates. Mortality data require large samples because child death events are rare. Our surveys collected baseline data from 2,894 women, of whom 2,187 ever gave birth, for a total of 9,536 birth histories. Following survey rounds had a similar size. This sample is not much smaller than the nationally representative samples collected by the DHS in Ghana, which ranged between 12,000 and 15,000 birth histories from 1988 to 2008. However, our sample is split into three groups of equal size, of which one is the MV sample, thus reducing statistical power considerably.

Mortality rates were considerably lower in MV areas before the intervention took place and the differences were statistically significant (Table 43). Infant and child mortality decreased in both MV and CV areas over the project period (Figure 48). The net DD effect is positive, meaning mortality decreased less rapidly in MV areas, though this is not statistically significant. Child mortality is one exception, which decreased more rapidly in MV areas, and the coefficient is statistically significant at 10%.

Figure 49 Infant, child and under-5 mortality in MV and CV areas

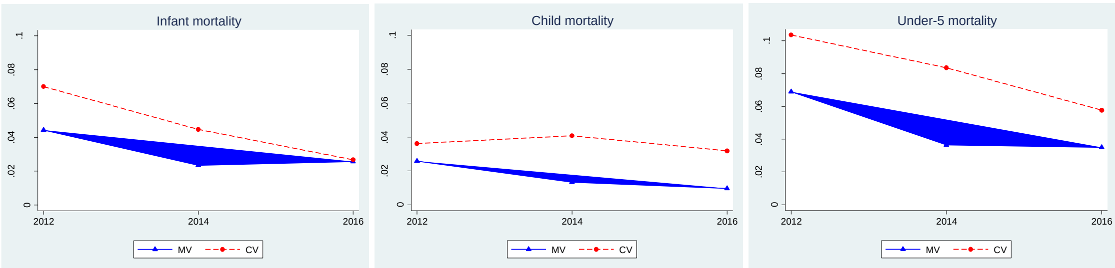


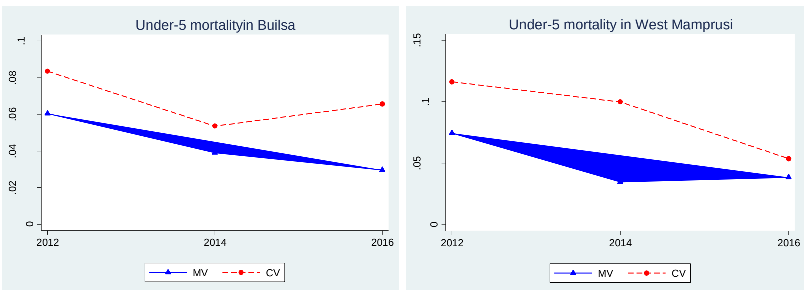
Table 43 Impact of MV on child mortality

	Baseline CV	Baseline diff. MV	DD impact midterm	DD impact endline
Neonatal mortality	38.52	-7.46	-28.22	6.12
		(0.540)	(0.139)	(0.675)
Post-neonatal mortality	31.42	-18.21**	19.55*	14.10
		(0.008)	(0.079)	(0.173)
Infant mortality	69.04	-25.67*	-8.67	20.22
		(0.097)	(0.711)	(0.285)
Child mortality	36.16	-10.38	-12.62	-15.77*
		(0.152)	(0.298)	(0.097)
Under-5 mortality	103.57	-34.66**	-20.86	4.12
		(0.030)	(0.389)	(0.842)

Note: Mortality rates calculated using synthetic cohort probability using SYNCMRATES. Coefficients estimated using IPW method. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster-adjusted standard errors

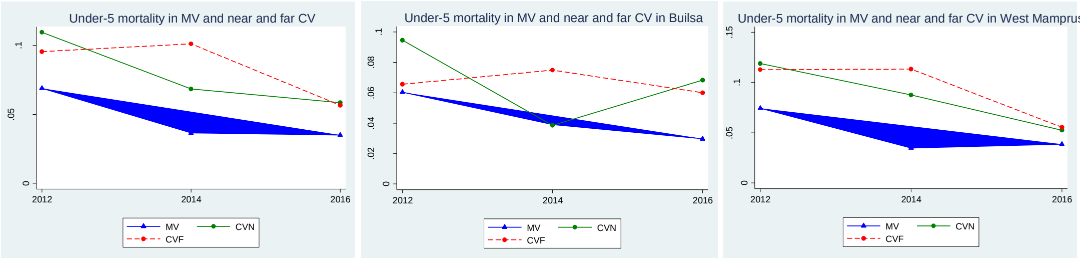
Patterns of change have been slightly different in the districts of Builsa and West Mamprusi (Figure 49). While more progress in mortality reduction was made in the MV communities in Builsa, the opposite is true for the control group communities, which saw more progress in mortality reduction in West Mamprusi.

Figure 50 Under-5 mortality in Builsa and West Mamprusi



Decreases in mortality rates were very similar in MV and near CV areas. The disaggregation of mortality rates by district and distance does not suggest the presence of spill-over effects. In Builsa mortality rates of near communities were very similar to those of MV areas at the midterm, but much higher at the endline, while in West Mamprusi the improvement in mortality rates in near communities was even larger than the improvement observed in the MV communities (Figure 50).

Figure 51 Under-5 mortality in near and far CV areas



We tried to address the low statistical power of our estimates by expanding the sample. The three surveys did not interview the same mothers. Birth histories are used to calculate mortality rates in the five years preceding the interview. There is therefore scope to add interviews of non-panel mothers whose births histories overlap. For example, the birth history of a mother interviewed at the endline, and who was not interviewed at the baseline, can be used to estimate mortality rates for the period preceding the baseline. Expanding the sample in this way, however, does not change estimates significantly. The baseline differences and the overall DD estimates turn out to be slightly smaller when using the expanded sample.

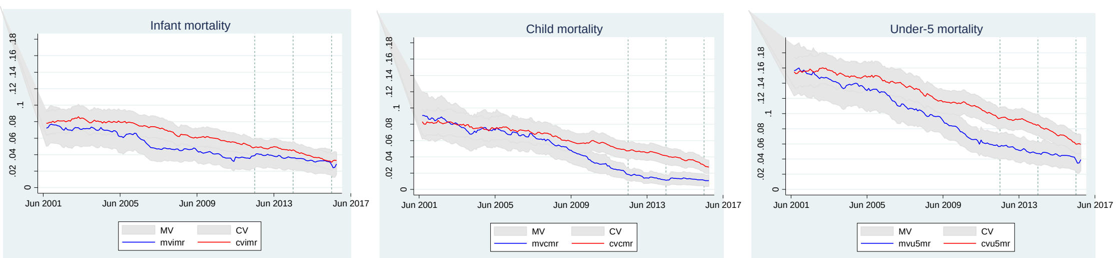
Table 44 Impact of MV on child mortality (expanded sample)

	Baseline CV	Baseline diff. MV	DD impact midterm	DD impact endline
Neonatal mortality	35.14	-10.19**	-22.22	9.49*
		(0.039)	(0.219)	(0.086)
Post-neonatal mortality	28.56	-12.37**	13.01	8.82
		(0.040)	(0.383)	(0.492)
Infant mortality	63.71	-22.26**	-9.21	13.31
		(0.039)	(0.767)	(0.311)
Child mortality	42.92	-18.66***	4.66	-0.77
		(0.000)	(0.529)	(0.869)
Under-5 mortality	103.89	39.48***	-5.22	16.49
		(0.000)	(0.868)	(0.288)

Note: Mortality rates calculated using synthetic cohort probability using SYNCMRATES. Coefficients estimated using IPW method. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster-adjusted standard errors

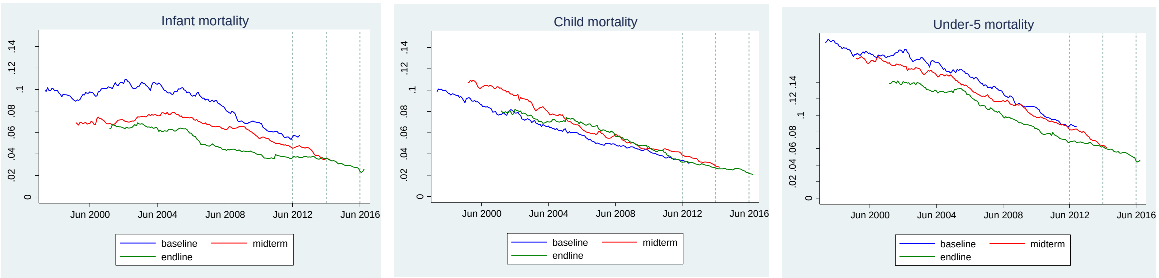
We used the expanded samples to calculate trends in mortality rates for 15 years before the endline (from 2001 to 2016). The results are shown in the charts of Figure 51. Mortality rates in MV and CV areas diverged well before the project and the differences in mortality rates between MV and CV areas narrowed after the intervention rather than increasing.

Figure 52 Mortality rates in MV and CV areas (2001-2016)



Finally, we plot mortality trends calculated from the three surveys separately and compare them on the same chart (Figure 52). Since the mortality rates are based on birth histories collected from the same mothers, they should largely overlap. However, the figure shows that mortality rates based on more recent survey rounds are consistently smaller. The difference is particularly large for the infant mortality rate, while the child mortality rate is similar across surveys. This might owe to age heaping. It is possible that, as time progresses, deaths defined as less than 12 months become defined as 12 months or more, thus moving deaths from the infant mortality rate to the child mortality rate over time.

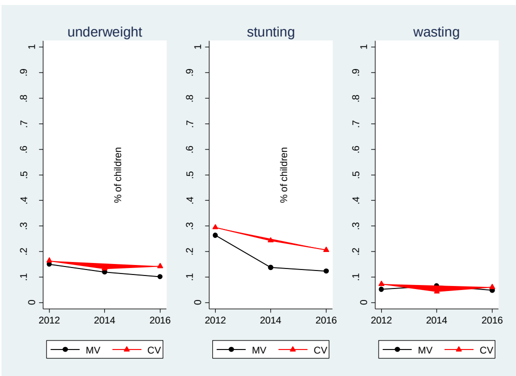
Figure 53 Overlapping mortality rates from different survey rounds



10.2. Impact on nutritional status of children

MV implemented a number of interventions expected to improve the nutritional status of children. These include provision of Vitamin A, de-worming and nutrition monitoring by CHWs. In addition, the programme promoted food production and food security and increased access to improved sources of drinking water and sanitation facilities. The programme acts on most determinants of undernutrition in the dimensions of food production, health and caring practices. The programme indeed had some impact on the nutritional status of children under five (Figure 53). Prevalence of underweight and stunting decreased at a faster rate in MV areas in comparison with CV areas.

Figure 54 Undernutrition prevalence rates in MV and CV areas



As already discussed when reporting the impact of the intervention on the MDGs, the size of the impact on underweight is not sufficiently large to reach statistical significance. But the reduction in the prevalence rates of stunting is large and is statistically significant (Table 45). Finally, wasting is getting worse, but this is simply a consequence of children's height improving at a faster rate than weight. The positive impact of MV on stunting is encouraging. Recall that stunting is an indicator of long-term undernutrition, that improvements in height last longer than improvements in height and that they reflect a general improvement in health conditions.

Table 45 Impact of MV on undernutrition prevalence rates

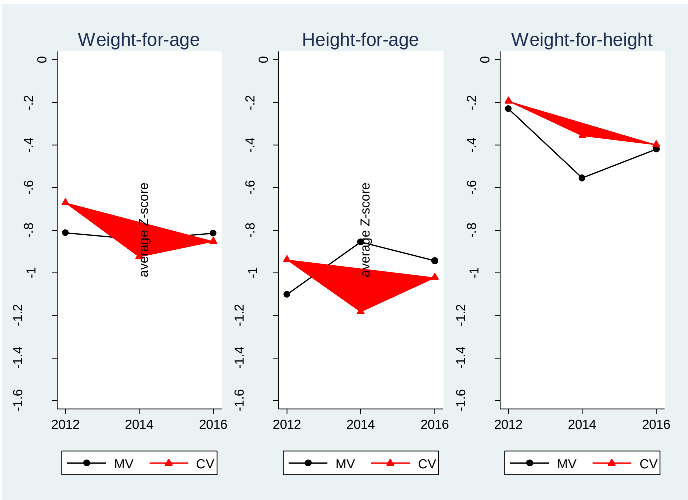
	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Underweight	16.43	-1.78	1.03	-2.15	-0.51
		(0.345)	(0.727)	(0.435)	(0.821)
Stunting	29.45	-3.45	-7.94**	-5.56	-6.85*

		(0.225)	(0.047)	(0.212)	(0.074)
Wasting	7.32	-2.22*	4.47**	1.90	3.22**
		(0.065)	(0.018)	(0.320)	(0.046)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

Prevalence rates are calculated as the proportion of children below 2 standard deviations from the reference mean (height or weight) for a specific age. These rates are commonly used in medical and public health practice and provide information on the general health conditions of a population. The cut-offs are set to identify children who are suffering or vulnerable to health conditions. Like all cut-offs, however, they hide as much as they reveal. Prevalence rates are blind to the distribution of outcomes and can remain unchanged despite significance changes in mean weight and height. For example, the project may improve weights of the most deprived children without letting them reaching the cut-off, or it could improve significantly the weights of children just above the cut-off. In both cases the project would be having an important impact that would go unnoticed by simply looking at prevalence rates. To shed more light on the impact of MV on undernutrition we also estimate the impact of the intervention on Z-scores that are used to estimate prevalence rates. The patters in Figure 54 show some erratic behaviours of Z-scores in both MV and CV areas, but also show that average Z-scores of weight-for-age and height-for-age are performing better in MV areas than CV areas.

Figure 55 Nutrition Z-scores in MV and CV areas



The improvements in average Z-scores are substantial in the case of weight-for-age, but particularly for height-for-age, and always statistically significant (Table 46).

Table 46 Impact of MV on nutrition Z-scores

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Average weight-for-age	-0.67	-0.13	0.26**	0.25**	0.25**
		(0.206)	(0.030)	(0.039)	(0.013)
Average height-for-age	-0.94	-0.15	0.51***	0.31**	0.42***
		(0.186)	(0.000)	(0.022)	(0.000)
Average weight-for-height	-0.19	-0.03	-0.14	0.05	-0.05
		(0.775)	(0.267)	(0.692)	(0.644)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

We calculated three indices recommended by the World Health Organization (WHO) to assess the quality of the diet of children under 24 months of age. Children are expected to consume at least four different types of food from a defined list of food categories, with a given frequency and in different combinations with fluids depending on whether they are breastfed or not. The diet diversity indices are the minimum dietary diversity (proportion of children between 6 and 23 months who eat food from at least 4 different categories), minimum meal frequency (the proportion of children between 6 and 23 months of age who consume solid foods and fluids a minimum number of times a day) and a minimum acceptable diet (the proportion of children who receive a minimum diversity of diet with minimum frequency). The project improved all dietary indicators between baseline and midterm but the difference between MV and CV areas decreased again at the baseline. The net effect is positive, and large, only for minimum dietary diversity. This is likely to be related to the widespread introduction of maize and beans in the family diet resulting from increased production of these specific crops.

Figure 56 Diet indices in MV and CV areas

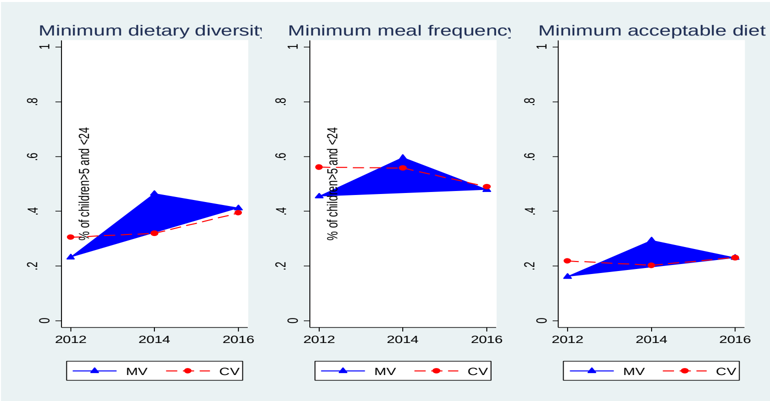


Table 47 Impact of MV on diet indicators

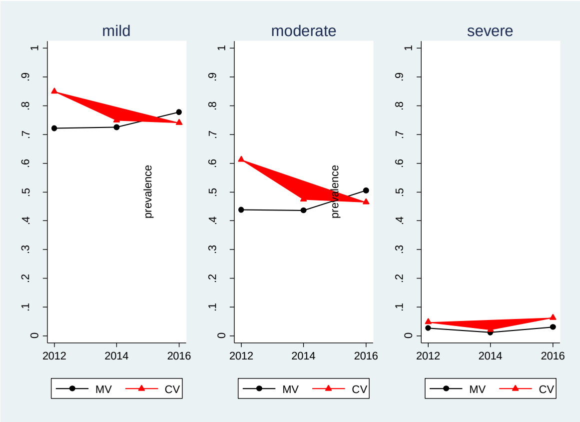
	Baseline CV	Baseline diff. MV	DD impact midterm 2014	DD impact endline 2016	Average DD impact
Minimum dietary diversity	30.51	-7.71	23.13**	8.36	17.08**
		(0.147)	(0.006)	(0.391)	(0.038)
Minimum meal frequency	56.12	-9.98*	10.60	4.94	8.42
		(0.091)	(0.193)	(0.622)	(0.268)
Minimum acceptable diet	21.88	-5.95	-3.76	1.61	1.12
		(0.154)	(0.520)	(0.833)	(0.833)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

10.3. Impact on prevalence of anaemia

The project also aimed at reducing prevalence of anaemia directly, through the provision of iron supplements, and indirectly through a reduction in morbidity rates. Trends in prevalence rates of moderate, mild and severe anaemia are shown in Figure 56. Note that the large baseline difference is a seasonal bias owing to the fact that the baseline blood tests were conducted during different seasons in the MV and CV areas. In particular, tests were conducted during the dry season in MV areas, when the prevalence of diarrhoea is moderate and so is iron depletion. The charts, however, do not show any progress in anaemia reduction in the midterm and endline survey, with the possible exception of severe cases. In fact, prevalence of anaemia appears to increase in MV areas in comparison with CV areas.

Figure 57 Prevalence of anaemia in MV and CV areas



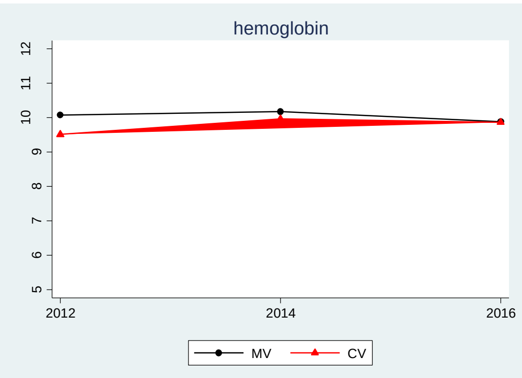
Statistical testing of changes in haemoglobin concentration and prevalence rates of anaemia confirm the visual impressions provided by Figure 57 (Table 48). There is a decrease in average haemoglobin concentration among children under five in MV areas and the prevalence rate of mild and moderate anaemia is increasing in comparison with in CV areas, and the difference is statistically significant.

Table 48 Impact of MV on children's anaemia

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Haemoglobin concentration	9.51	0.55**	-0.17	-0.56**	-0.40**
		(0.001)	(0.502)	(0.001)	(0.041)
Mild anaemia	84.97	-12.42**	7.08	16.87***	13.03**
		(0.002)	(0.353)	(0.000)	(0.014)
Moderate anaemia	61.27	-17.10**	11.29	22.65***	17.09**
		(0.001)	(0.152)	(0.000)	(0.005)
Severe anaemia	4.83	-2.09	1.87	-0.96	0.12
		(0.202)	(0.428)	(0.728)	(0.957)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

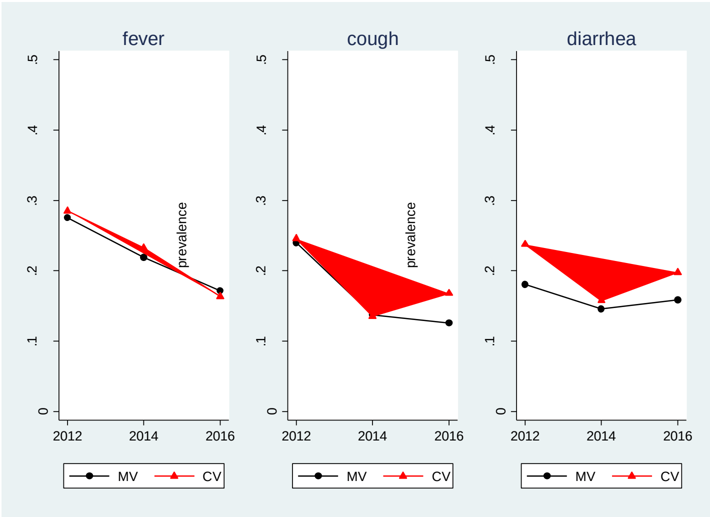
Figure 58 Haemoglobin concentration in MV and CV areas



10.4. Impact on malaria and other common diseases

During the adult surveys, mothers reported on the occurrence of episodes of fever, cough and diarrhoea for all children under five in the household. Trends in prevalence rates are fairly similar across MV and CV areas (Figure 58), with the possible exception of cough in the last survey round.

Figure 59 Prevalence of cough, fever and diarrhoea among children under five



None of the prevalence rates shows a sizeable reduction (except cough at endline) and average impacts are not statistically significant (Table 49). The project does not appear to have reduced the incidence of most common symptoms of diseases affecting the population in the study area.

Table 49 Impact of MV on fever, cough and diarrhoea of children under five

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Fever during past 2 weeks	28.52	-1.24	-0.35	2.21	0.96
		(0.674)	(0.935)	(0.567)	(0.789)
Illness with cough during past 2 weeks	24.58	-0.52	0.64	-3.41	-1.42
		(0.870)	(0.878)	(0.375)	(0.695)
Diarrhoea during past 2 weeks	23.76	-5.70**	4.31	1.98	3.30
		(0.028)	(0.202)	(0.636)	(0.314)

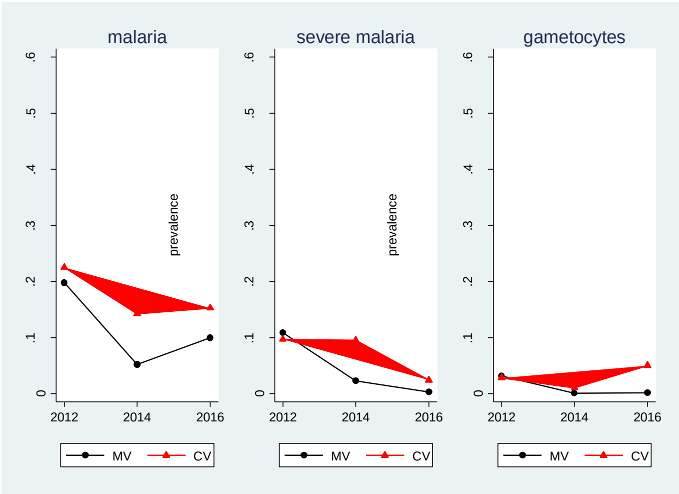
Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

The results of the project impact on the incidence of malaria are more encouraging. Malaria infection was assessed by the survey through blood tests and thick and thin smears from children under five. The slides with the blood samples were examined twice for the presence of the malaria parasite (the trophozoites stage). If both tests were positive, the child was coded as infected by the parasite. If only one of the tests was positive, the slide would be tested a third time and if positive the child would be coded as infected by the parasite. Blood samples of children coded as positively infected by the malaria parasite would then be examined to assess the density of the parasite to assess the severity of malaria.

Parasite infection is common and low levels of infection are not harmful. The severity of malaria is estimated by counting the number of parasites per white blood cells (if no parasites are found in the first 200 white cells, a further 300 cells are examined). The results are then expressed in terms of parasites per microliter of blood. Finally, the slides are examined for gametocytes, which are the sexual cells of malaria. If one gametocyte is discovered, then a count is conducted on the number of gametocytes. This count is done for every 200 or 500 leukocytes, whichever number was used to count the number of parasites. The density of gametocytes is an indicator of how easily the subject can transfer the malaria to mosquitos, which can then transfer it to other humans.

The charts in Figure 59 seem to show some project impact on malaria incidence. Malaria prevalence is defined as the percentage of children coded as infected by the malaria parasite, while severe incidence was defined as the proportion of children with parasitic infection above the baseline median parasitic infection of children coded as infected. The presence of gametocytes also appears to be larger in CV areas than in MV areas.

Figure 60 Malaria prevalence in MV and CV areas



The intervention had a negative impact on overall malaria incidence but the size of the impact is not sufficiently large to reach statistical significance (Table 50). We find, however, a statistical significant impact when we consider only cases of severe malaria, defined as the proportion of children infected by a larger than average level of parasitic infection. Recall that malaria parasitic infection is common and that many cases of infection caused by a small amount of parasites do not result in health conditions. The project also has a negative impact on the number of gametocytes. Recall that gametocytes are responsible for the transmission of the infection from one subject to the other, so a reduction in their number suggests a potential lower contagion of the parasite infection to other subjects. In other words, there is some evidence from blood tests that the project reduced malaria parasitic infection and its potential to spread.

Table 50 Impact of MV on malaria infection

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Malaria incidence	22.51	-2.88	-4.50	-4.47	-5.53
		(0.471)	(0.333)	(0.345)	(0.196)
Severe malaria incidence	9.69	0.97	-7.46**	-3.50	-5.71**
		(0.704)	(0.026)	(0.196)	(0.050)
Presence of gametocytes	2.82	0.40	-0.70	-5.78**	-3.70**
		(0.760)	(0.575)	(0.009)	(0.012)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

The project aimed at reducing the prevalence of malaria through preventative measures and information campaigns. Households were distributed insecticide-treated bednets and parents were instructed on their use and on the perils of mosquitos' bites by CHWs. Populations in the study area have a number of misconceptions regarding the causes of malaria transmission, such as that malaria can be caught through excessive exposure to the sun, eating sweets or witchcraft. We tested the extent of belief in these misconceptions and also built an index running from 0 to 4 (where 4 means maximum misconception). The project does not have an impact on any of these erroneous beliefs, though it increased the number of people believing malaria could be transmitted by mosquito bites (Table 51). This impact is statistically significant, but it should be noted that only 3.5% of the MV sample held an erroneous belief in this regard at the baseline. Other beliefs were not affected by the project and it should be noted that the extent of these beliefs is still very high.

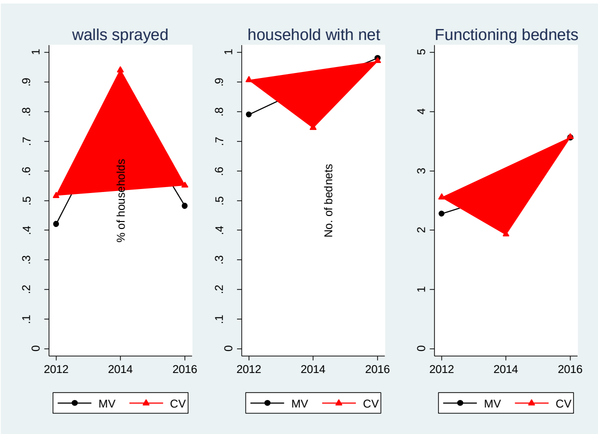
Table 51 Impact of MV on adults' knowledge of malaria

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
A person can get malaria from standing in the sun too long	78.49	-2.06	-0.09	2.66	1.21
		(0.463)	(0.979)	(0.407)	(0.702)
A person can get malaria from a mosquito bite	97.95	-1.45*	2.30**	1.53	1.93**
		(0.063)	(0.013)	(0.107)	(0.025)
A person can get malaria from eating sweets	60.10	-6.37	1.08	6.16	3.50
		(0.64)	(0.792)	(0.127)	(0.341)
A person can get malaria from witchcraft	47.49	-2.65	2.82	-3.40	-0.16
		(0.528)	(0.582)	(0.485)	(0.969)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

The project made considerable efforts in distributing mosquito bednets and promoting their use. The section on the impact on the MDGs has already shown a dramatic impact on the proportion of children sleeping under mosquito nets. Here we look at three additional indicators of the use of mosquito bednets and other forms of malaria prevention: whether walls had been sprayed with insecticides, whether the household has any mosquito nets (observed, not reported) and the number of functioning nets. The project did not promote spraying of walls, and the prevalence of spraying followed similar patterns in MV and CV areas presumably led by other organisations or by the government (Figure 60). The number of households with mosquito nets and the overall number of functioning nets increased. But so did the same indicators in the CV areas. However, the CV areas witnessed a reduction in the availability of nets that was not recorded in MV areas.

Figure 61 Malaria prevention in MV and CV areas



The patterns of net availability over time have some interesting characteristics. First, the baseline difference is likely the result of the seasonal bias already observed in the measurement of anaemia. Since the surveys were conducted in the dry season in MV areas the reporting of mosquito net was lower because traditionally mosquito nets are obtained and used during the rainy season, when mosquitos pose a much higher threat. As a result, the increase in the availability of bednets appears more dramatic than it actually was. On the other hand, there was a considerable increase in the availability of mosquito nets in CV areas at the endline, probably resulting from similar interventions being implemented by other organisations or by the government. The difference in net availability was largest at the midterm, when CV areas showed a large decrease in availability. In sum, the impact of the project on the use of mosquito nets appears substantial and is statistically significant (Table 52). However, it should be noted that, because of seasonality bias at baseline data collection and of particular patterns in CV areas, it may appear larger than it actually was.

Table 52 Impact of MV on use of mosquito bednets

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact

Walls sprayed in past 12 months	51.60	-9.12**	-1.46	3.17	0.87
		(0.044)	(0.880)	(0.503)	(0.878)
Household have any mosquito nets	90.74	-11.74**	26.49***	12.51***	19.51***
		(0.003)	(0.000)	(0.000)	(0.000)
Number of functioning nets	2.56	-0.28	0.96***	0.26	0.61**
		(0.129)	(0.000)	(0.269)	(0.001)

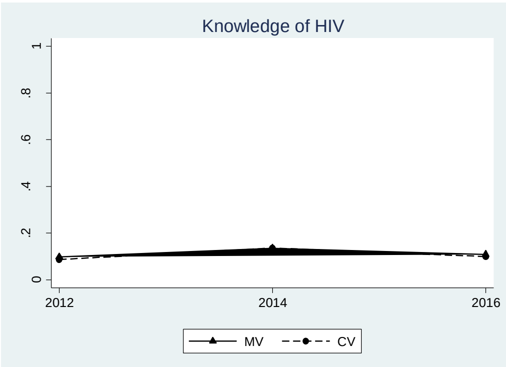
Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

10.5. Impact on knowledge, attitudes and practices

We report here the responses to questions to adults of both sexes aged 15-49 aimed at assessing the level of knowledge of best health practices. The surveys included questions on knowledge of the causes of HIV (also used to monitor the MDGs), breastfeeding and knowledge of best hand washing practices.

Following MDG practices we calculated the proportion of adults 15-49 who correctly answered eight questions about obvious causes of HIV infection transmission. The proportion of adults with correct knowledge of HIV measures of this type is extremely low in the study area and did not improve after the project (Figure 61).

Figure 62 Knowledge of HIV in MV and CV areas



Adults were asked about all of the occasions on which it is important to wash hands (Table 53). Enumerators were instructed not to read the possible answers and to report all the answers provided. The project did not change people’s attitudes to hand washing, with just one exception. Less than 60% of adults reported that it was important to wash hands after defecation at baseline but this percentage increased dramatically in the project areas, and the difference is statistically significant. On the other hand, the project did not seem to have an impact on the relatively low proportions of individuals who believe hands should be washed before cooking, before feeding a child and after cleaning a toilet or a potty.

Table 53 Impact of MV on hand-washing practices

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Important to wash hands before eating	95.71	-1.93	1.21	3.69	2.40
		(0.276)	(0.438)	(0.153)	(0.156)
Important to wash hands before breastfeeding or feeding a child	25.58	6.23	-2.70	1.35	-0.76
		(0.205)	(0.609)	(0.859)	(0.893)
Important to wash hands before cooking or preparing food	49.17	4.17	2.57	3.36	2.95
		(0.359)	(0.595)	(0.626)	(0.544)
Important to wash hands after defecation or urination	57.54	-6.79	20.30**	15.94**	18.18**
		(0.244)	(0.001)	(0.010)	(0.003)
Important to wash hands after cleaning a child who has defecated	30.04	10.08*	-1.77	-1.98	-1.86
		(0.062)	(0.766)	(0.817)	(0.783)
Important to wash hands after cleaning toilet or potty	26.23	2.42	5.54	10.11	7.75

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

We further analysed the adoption of two health practices that are immediately relevant to child health: breastfeeding and vaccinations. Breastfeeding is universally practised in the area and the project could not increase the number of children ever breastfed (Figure 62). Early initiation of breastfeeding within the first hour from birth increased considerably and at the same rate in MV and CV areas. The proportion of women exclusively breastfeeding in the first three days and the first six months without administration of other fluids or solid is relatively large in the area and the project preserved these practices more than in CV areas. The impact on use of mothers' milk in the first three days after birth and on exclusive breastfeeding is large and statistically significant (Table 54)

Figure 63 Prevalence of breastfeeding practices

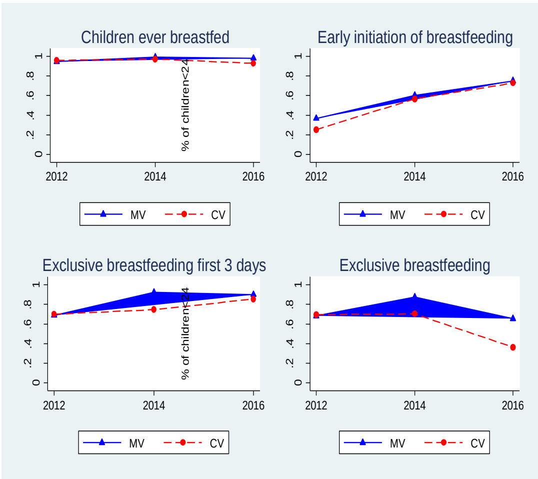


Table 54 Impact of MV on breastfeeding practices

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Child ever breastfed	95.98	-1.23	3.71	6.85	4.97
		(0.674)	(0.241)	(0.113)	(0.152)
Breastfed within 1 hour of birth	25.17	11.48**	-6.74	-8.35	-6.99
		(0.018)	(0.455)	(0.298)	(0.234)
No fluids other than milk during first 3 days	70.02	-1.07	18.38**	6.47	13.10**
		(0.858)	(0.002)	(0.395)	(0.043)
Exclusive breastfeeding first 6 months	69.65	-1.37	16.85**	29.28***	22.50**
		(0.840)	(0.046)	(0.000)	(0.004)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

The project increased the prevalence of vaccinations among children under two for three major vaccines (BCG, DPT and measles), though did not change polio vaccination rates. All the differences are statistically significant (Table 55). The project did not appear to increase children's intakes of vitamin A and deworming tablets. This latter finding is rather odd considering the project's early emphasis on the 'quick wins', which included the supplementation of vitamin A and deworming.

Table 55 Children's vaccination rates and supplements

	Baseline CV	Baseline diff. MV	DD impact midterm 2014	DD impact endline 2016	Average DD impact
Vaccination card	65.7	12.1***	10.0***	7.8***	9.2***
		(0.001)	(0.000)	(0.000)	(0.000)
BCG	81.8	3.2	5.1***	2.7*	4.0***
		(0.236)	(0.004)	(0.054)	(0.004)
Polio	43.3	-2.6	-2.8	-4.4	-3.4
		(0.553)	(0.451)	(0.354)	(0.365)

DPT	66.5	5.5	8.1***	5.8**	7.1***
		(0.244)	(0.008)	(0.017)	(0.003)
Measles	69.9	1.1	4.9*	5.0**	5.1**
		(0.722)	(0.051)	(0.027)	(0.010)
Vitamin A	69.5	-0.4	-9.5**	-2.7	-6.4*
		(0.939)	(0.014)	(0.531)	(0.068)
Deworming	38.5	1.7	-2.7	8.2	2.7
		(0.726)	(0.554)	(0.135)	(0.493)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

11. Impact of MV on education

MVP aims to achieve 'enhanced access to quality primary education' through five output areas:[15]

1. Improving education quality
2. Increasing primary school enrolment
3. Increasing participation in secondary education
4. Improving gender parity
5. Engaging communities in education

The MVP aimed to achieve the anticipated results through a range of activities delivered across the education sector. In the first year of operation, project staff conducted several needs assessments with the communities, PTAs, school management committees (SMCs) and district education directorates. The meetings revealed the scale and variety of problems faced by the education system in the north: inadequate buildings and teaching materials, teacher absenteeism, poor teacher qualifications, high teacher turnover, language barriers to learning, economic and social constraints to school attendance such as long distances to school, absence of toilets for girls and the low value parents place on schooling.

The project devised an overall strategy to tackle these problems with the main goal of increasing school attendance. The strategy was based on delivering **activities** within three main pillars: 1) improving school quality; 2) sensitising communities and parents; and 3) enrolling more girls in school. Additional interventions aimed at bringing more children to school and monitoring children's/students' learning were attempted, but on a much smaller scale.

- **Schooling quality:** It was thought that one of the main factors behind low school attendance was the poor quality of instruction. This in turn was the result of poor school infrastructure and poor teaching (including the intimidation of children). Hence, the project invests heavily in the construction and rehabilitation of classrooms, school toilets and playgrounds, and refurbishes schools with sporting equipment, teaching materials, books and computers. In order to increase the *quality of teaching*, the project builds teacher quarters and provides other incentives for teachers to live in the communities. The project trains teachers on teaching methods and provides salary top-ups to staff of the Ghana Education Service (GES) to supervise teachers' work.
- **Community:** The aim of the community sensitisation work is to strengthen communities' understanding of the role they can play in advancing children's education (e.g. by ensuring children get to school on time, holding schools, head teachers and teachers accountable for children's performance, the school holding community members accountable for their responsibilities to children, etc.). The project hires and trains community

[15] 2012, Annual Report on the Millennium Villages Project in Northern Ghana, p. 10.

education workers (CEWs) with the goal that they will hold meetings and workshops with the communities, PTAs and SMCs to sensitise parents about the benefits of education. In addition, CEWs visit families of children not attending school and families of children who have dropped out of school to get more children in school.

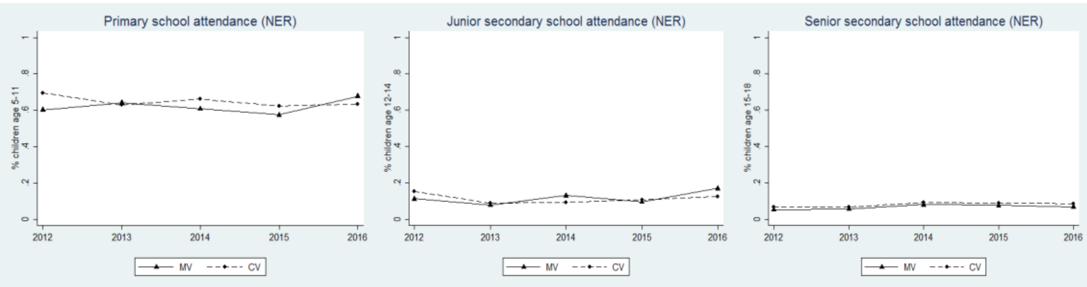
- **Gender parity:** In order to boost girls' school attendance, MVP implemented a varied set of initiatives, including school toilets for girls, delivery of sanitary pads to prevent absence from school during menstruation, community and parents' sensitisation on the benefits of girls' schooling and scholarships for girls attending senior secondary schools.

In addition to these broad packages of initiatives, the project also tried to increase school attendance directly by supporting the provision of school meals[16] and establishing a real-time monitoring system in schools to improve learning. This was performed by CEWs using mobile technology. CEWs would assess students' reading skills on a regular basis to inform project staff and education authorities about progress being made and establish areas where remedial education was needed.

We assess impact on school attendance by calculating net attendance ratios. The attendance ratio calculates the proportion of children of a specific age who report having attended, over the previous year, the appropriate school for their age. For example, the net attendance rate in primary consists of the proportion of children aged 6-11 who attended primary school in the year preceding the interview. Net attendance ratios for junior secondary and senior secondary are calculated in a similar way. Since late entrants (children starting school at an older age than 6) and returning pupils (children returning to school after long breaks out of school) are common, we also calculate an attendance rate for primary for all children aged 5 to 18, and an attendance rate of any school level for the same age group.

The charts in Figure 63 do not show any obvious impact of the intervention on attendance ratios. However, the DD estimates provide a more positive picture (Table 56). The project had a positive impact on all attendance indicators except senior secondary schools. The impact on primary attendance was of nearly 8% on average and was largest in the first and the last year of intervention. The impact on attendance of junior secondary is nearly 6% and fails to achieve statistical significance. Attendance of any school by all children aged 5-18 also increases by 5%, reflecting larger attendance of school by children outside the school specific age range. Much of the project efforts were focused on increasing attendance in primary schools, but completion of primary increases the probability of attending junior secondary and senior secondary and therefore we would expect also attendance of higher grades to increase over time. A small increase in junior secondary occurred but there was no change in senior secondary, which is currently attended by a small fraction of the student population. The changes observed in junior and senior secondary school were not statistically significant.

Figure 64 Net attendance ratios in primary, junior and senior secondary school



[16] The provision of school meals was under the existing school feeding programme. It is not a programme specific to MVP, but rather is supported by the project.

Figure 65 School attendance of any grade

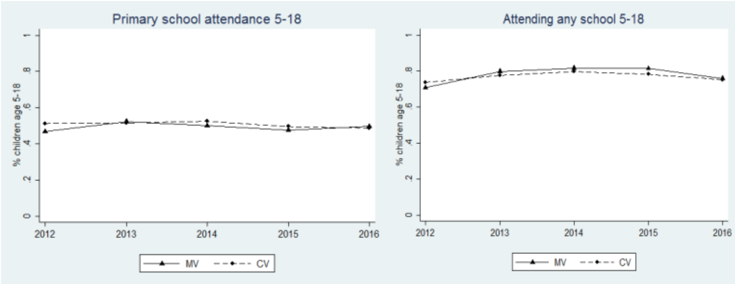


Table 56 Impact of MV on school attendance

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2014	DD impact 2015	DD impact 2016	Mean impact	Sample size
Primary education	0.696	-0.093* (0.056)	0.096*** (0.008)	0.043 (0.265)	0.035 (0.338)	0.135*** (0.000)	0.077** (0.017)	13,994
Junior secondary school	0.154	-0.038 (0.171)	0.027 (0.408)	0.081* (0.060)	0.028 (0.459)	0.089* (0.087)	0.057 (0.119)	5,671
Senior secondary school	0.069	-0.016 (0.387)	0.002 (0.913)	0.001 (0.979)	0.005 (0.824)	-0.004 (0.865)	0.001 (0.969)	7,147
Primary attendance (children 5-18)	0.513	-0.045 (0.175)	0.047* (0.052)	0.017 (0.474)	0.020 (0.411)	0.054* (0.059)	0.035 (0.105)	29,346
Attendance of any school (children 5-18)	0.708	-0.030 (0.481)	0.057** (0.017)	0.055** (0.020)	0.061** (0.010)	0.040 (0.131)	0.053** (0.016)	29,346

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

Our study measured learning in school by administering a set of tests to children. All children were administered three cognitive tests: Raven's matrices and forward and backward digit spans. The selected cognitive tests measure different dimensions of 'intelligence' and capture genetic as well as acquired skills. Children who are physically and intellectually stimulated at a young age tend to perform better at these tests. Simple (8-question) maths and English tests were administered to children aged 6-11 who ever attended primary, and advanced (and much longer) maths and English tests were administered to children older than 11 who ever attended junior secondary school.

The project did not improve children's cognitive skills (Table 57). Oddly, it appears to have had a negative impact on the backward digit span test. The negative effect is consistent across the midterm and the endline assessment and of similar size. In a digit span test, the subject is requested to repeat a sequence of random numbers. The backward digit span test is more challenging than the forward digit span test as it requires the subject to repeat the series of numbers in reverse. The test measures the size of short-term memory, which can be affected by learning practice (for example practising music increases short-term memory) or by factors related to attention, such as a proper diet and micronutrient intake (malnourished and anaemic children tend to perform more poorly).

Table 57 Impact of MV on test scores

	Baseline CV	Baseline diff. MV	DD impact midterm	DD impact endline	Average DD impact	Sample size

Raven matrices test	0.010	-0.007 (0.956)	0.236* (0.080)	-0.197 (0.321)	-0.002 (0.991)	10,602
Forward digit span	0.015	-0.036 (0.681)	-0.156 (0.130)	-0.027 (0.818)	-0.086 (0.380)	10,508
Backward digit span	-0.007	0.078 (0.403)	-0.230* (0.099)	-0.263** (0.040)	-0.248** (0.033)	10,503
Easy maths	0.003	0.009 (0.911)	0.099 (0.403)	-0.330** (0.021)	-0.147 (0.219)	5,956
Easy English	0.053	-0.100 (0.297)	0.066 (0.542)	-0.386*** (0.002)	-0.190* (0.055)	5,581
Advanced maths	0.029	-0.099 (0.448)	0.342* (0.067)	-0.479*** (0.005)	-0.158 (0.290)	1,674
Advanced English	0.038	-0.122 (0.279)	0.567*** (0.002)	-0.299 (0.127)	0.035 (0.827)	1,683

Note: The dependent variable is reported in standard deviations of the baseline average combined in the project and control groups. Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

The impact on most test scores is negative with the exception of the advanced English test. The negative effects are not too large (under 0.2 standard deviations) and only the effect on the easy English test is statistically significant. Since easy tests were administered to children who ever attended primary school, while advanced tests were administered to children who ever attended junior secondary, this result fits well with a story of an increase in school attendance in MV areas by children of poorer backgrounds and with no previous education. The increase in school attendance in MV areas brought to school children who had never attended school or who were more disadvantaged to start with, and who therefore tend to perform more poorly in tests. Indirect evidence of this was obtained by looking at the impact of the project on a panel of children tested both at the baseline and at the endline and whose test scores did not change or improve, though the change was never statistically significant (Table 58). Alternative explanations should refer to the quality of teaching in the MV areas and could point to factors such as room overcrowding or the employment of less qualified teacher by the project.

Table 58 Impact of MV on test scores (only panel children)

	Baseline CV	Baseline diff. MV	DD impact midterm	DD impact endline	Average DD impact	Sample size
Raven matrices test	-0.129	-0.020 (0.869)	0.259** (0.014)	-0.277* (0.052)	-0.008 (0.917)	1,994
Forward digit span	-0.170	-0.008 (0.943)	-0.092 (0.444)	0.064 (0.653)	-0.014 (0.899)	1,910
Backward digit span	-0.238	0.138 (0.147)	-0.090 (0.589)	-0.044 (0.769)	-0.067 (0.589)	1,904
Easy maths	-0.278	0.214 (0.118)	0.139 (0.400)	-0.143 (0.412)	-0.001 (0.997)	728
Easy English	-0.150	-0.076 (0.592)	0.141 (0.631)	0.030 (0.902)	0.090 (0.683)	530
Advanced maths	0.176	-0.454 (0.319)	0.163 (0.668)	-0.272 (0.553)	-0.059 (0.825)	36
Advanced English	-0.160	-0.443 (0.156)	-0.091 (0.895)	0.755 (0.224)	0.351 (0.338)	38

Note: The dependent variable is reported in standard deviations of the baseline average combined in the project and control groups. Coefficients are fixed effects DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

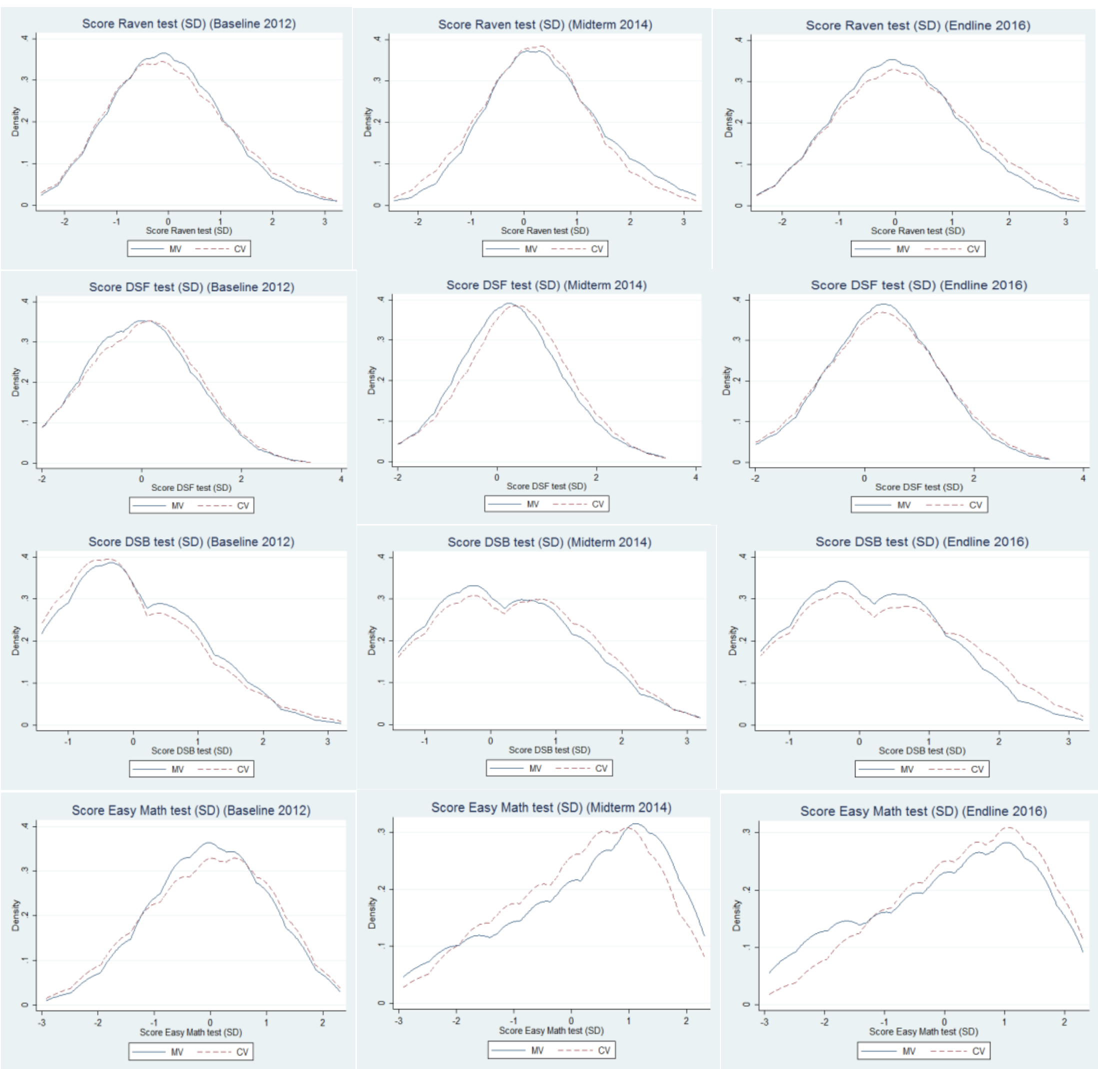
Our survey also asked parents and children about their wage expectations. Respondents were asked to report the wages they thought would be paid to a person with primary school qualifications and to a person with a secondary school degree. Interestingly, at baseline, expectations were much lower in MV areas compared with CV areas, possibly reflecting lower prevailing wages. At baseline parents' and children's expectations were very similar to each other, though children's perceptions of wages were on average higher. The project did not change children's wage expectations. However, parents' wage expectations increased substantially both for individuals holding a primary degree and those with a secondary school degree (Table 59). In the economics of education theory, it is often assumed that people make schooling decisions based on, among other thing, the expected wage for different schooling levels. In such a framework, an increase in expected wages should determine an increase in schooling. It remains to be explained whether parents' change in expectations is the result of a better understanding of the benefits of schooling produced by the project's social mobilisation work or is of a true increase in wages in the area. From an economics perspective, however, we interpret changes in parents' wage expectations as positive predictors of a higher probability of sending children to school.

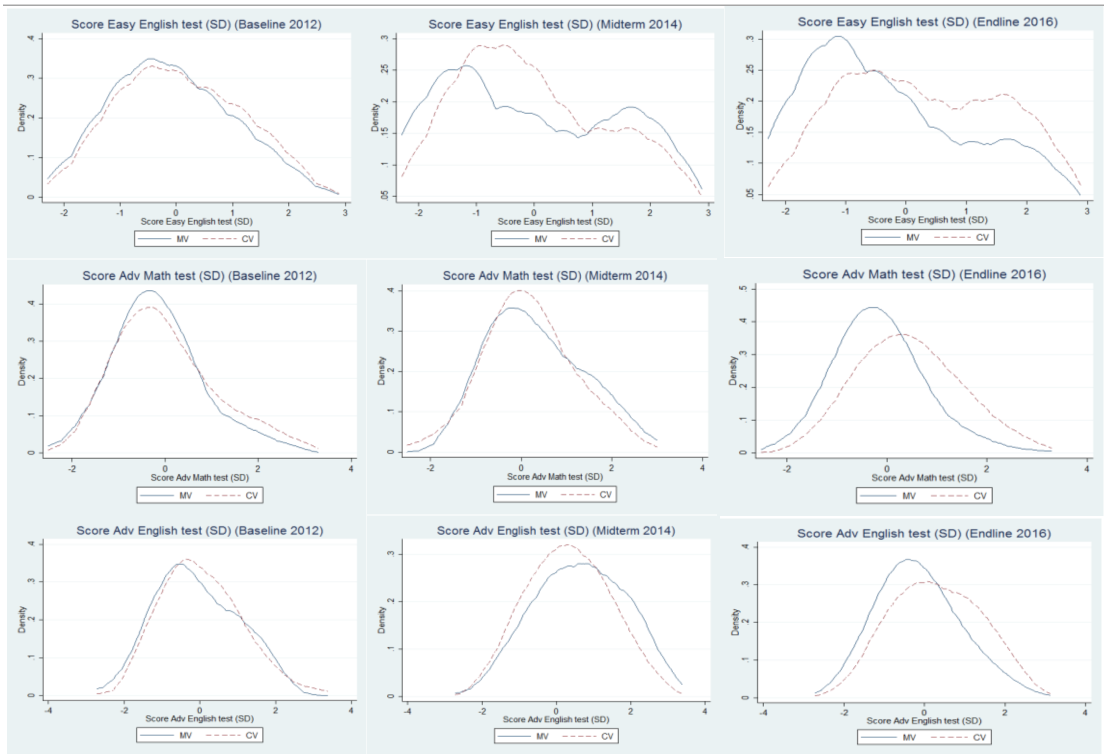
Table 59 Impact of MV on parents' and children's wage expectations

	Baseline CV	Baseline diff. MV	DD impact midterm	DD impacts endline	Average DD impact
Children wage expectations (primary school)	7.89	-3.44***	1.96	-12.30	-6.20
		(0.000)	(0.332)	(0.270)	(0.334)
Children wage expectations (secondary school)	17.06	-10.80***	11.99**	-5.44	1.98
		(0.000)	(0.037)	(0.740)	(0.826)
Parents wage expectations (primary school)	7.07	-2.20***	1.60	1.85**	1.73**
		(0.000)	(0.018)	(0.012)	(0.006)
Parents wage expectations (secondary school)	13.99	-7.22***	7.98	5.53**	6.75***
		(0.000)	(0.000)	(0.001)	(0.000)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

Figure 66 Density distributions of test scores





12. Other miscellaneous impacts

12.1. Water and energy security

The survey asked respondents whether their household had enough drinking water and fuel for cooking over the previous 12 months (Table 60). The impact was negative for both, meaning a reduction in household insecurity, but the effects were not statistically significant.

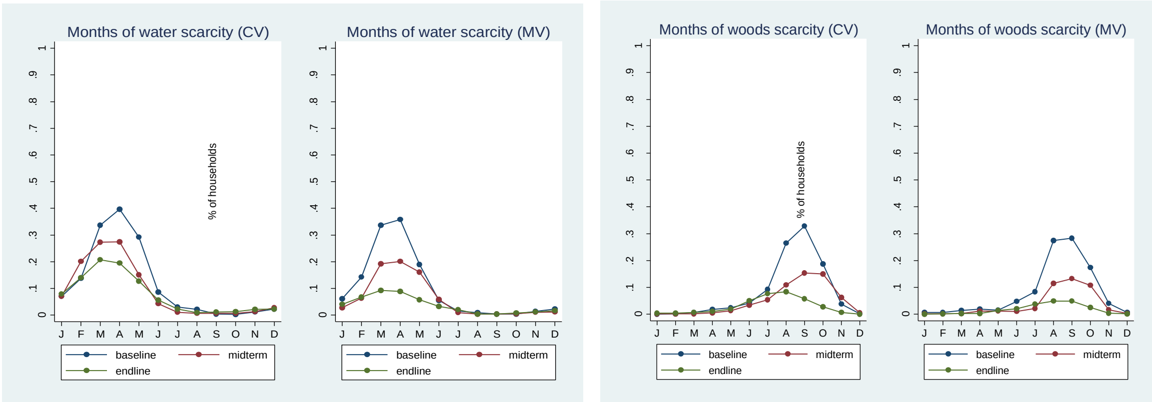
Table 60 Impact of MV on water and energy security

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2015	DD average impact
Not enough water in past 12 months	47.58	-4.73	-4.84	-8.60	-6.70
		(0.540)	(0.392)	(0.191)	(0.240)
Not enough woods for fuel in past 12 months	37.52	-3.50	-1.80	-0.85	-1.30
		(0.542)	(0.763)	(0.879)	(0.805)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

Households were asked to report the months in which they experienced most scarcity of water and woods for cooking (Figure 66). Seasonal water stress is highest in the dry season between March and May, while scarcity of wood is most common between August and October and of food in the months between May and August. Both seasonal water and woods scarcity appear to decrease over the project period. The change is nearly identical in MV and CV areas in the case of woods, while in the case of water it appears to decrease more in CV areas.

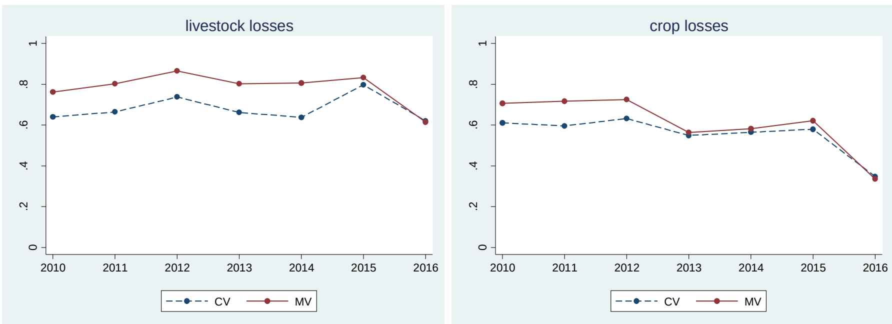
Figure 67 Seasonal water and energy insecurity



12.2. Vulnerability and coping strategies

The surveys collected data on the occurrence of significant livestock and crop losses. The baseline survey further collected data on shocks suffered in the two years preceding the survey (Figure 67) The percentage of households reporting livestock and crop shocks is very large in both MV and CV areas. A significantly larger number of households reported livestock and crop losses at the baseline and for the two years before the baseline. After the project started, the differences decreased and the prevalence of livestock and crop shocks appears to be identical at the time of the endline survey.

Figure 68 Incidence of livestock and crop losses in MV and CV areas



The impact of MV on the occurrence of both shocks is negative but is statistically significant for livestock losses and only in the last two survey rounds (Table 61). It is difficult to say to what extent these changes represent changes in the occurrence of shocks or changes in people's perceptions of the importance of the shocks.

Table 61 Impact of MV on livestock and crop shocks

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2014	DD impact 2015	DD impact 2016	Mean impact
Livestock losses	67.53	12.32**	0.53	2.79	-10.73**	-17.32**	-6.14
		(0.021)	(0.919)	(0.610)	(0.025)	(0.007)	(0.161)
Crop losses	61.55	9.54**	-6.05	-10.13	-5.66	-13.09	-8.71
		(0.038)	(0.383)	(0.175)	(0.437)	(0.102)	(0.153)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

The adult survey also investigated how people would behave when facing adversities. In particular, male and female adults were asked whom they would ask for financial support and shelter if in need. The questions are entirely hypothetical and a series of possible 'helpers' are suggested: family members, neighbours, government officers etc. The project appears to have a consistent impact on the proportion of people who would be confident to ask for help from neighbours and an NGO for financial support and shelter (Table 62). In the case of shelter, the confidence seems to extend to other people in the village, whom the respondent does not know so well. The project does not increase confidence in relying on family members, government officers and people in the village whom the respondent does not know (the latter has a statistically significant negative impact). These data seem to support the fact that in MV areas people are more confident to ask for financial help and shelter from neighbours and NGOs.

Table 62 Impact of MV on coping strategies

	Baseline CV	Baseline diff. MV	Comp. change 2014	Comp. change 2016	Average comp. change	Sample size
Ask money family	0.925	0.009	-0.019	0.035	0.007	12,694
		(0.623)	(0.362)	(0.138)	(0.739)	
Ask money neighbour	0.773	-0.034	0.079*	0.206***	0.140***	12,635
		(0.232)	(0.073)	(0.000)	(0.003)	
Ask money people in the village	0.213	-0.039	0.061	0.092	0.076	12,436
I do not know so well		(0.372)	(0.242)	(0.113)	(0.118)	
Ask money people in the village	0.139	0.027	-0.139***	-0.021	-0.082*	12,341
I do not know		(0.523)	(0.005)	(0.638)	(0.057)	
Ask money NGO	0.486	0.049	0.102	0.112*	0.107*	12,361
		(0.414)	(0.162)	(0.072)	(0.090)	
Ask money government officials	0.484	0.115**	-0.033	-0.057	-0.044	12,305
		(0.018)	(0.615)	(0.303)	(0.435)	
Ask shelter family	0.948	0.012	-0.017	0.017	-0.001	12,723
		(0.261)	(0.124)	(0.245)	(0.939)	
Ask shelter neighbour	0.806	-0.033	0.103**	0.111**	0.106***	12,667
		(0.304)	(0.020)	(0.014)	(0.007)	
Ask shelter people in the village	0.226	-0.054	0.165***	0.115**	0.141***	12,428
I do not know so well		(0.247)	(0.004)	(0.042)	(0.003)	
Ask shelter people in the village	0.146	0.014	-0.117**	0.020	-0.050	12,344
I do not know		(0.724)	(0.028)	(0.681)	(0.243)	
Ask shelter NGO	0.534	-0.001	0.196***	0.163***	0.179***	12,354
		(0.990)	(0.006)	(0.006)	(0.003)	
Ask shelter government officials	0.538	0.021	0.114*	0.070	0.092	12,289
		(0.666)	(0.071)	(0.254)	(0.114)	

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

12.3. Impact on migration

Different hypotheses can be formulated with respect to whether the project should increase or decrease migration from the intervention areas. Ultimately, the impact of the intervention on migration is an empirical question, albeit a very important one because it can potentially affect the comparison between project and control areas. In this section we make an attempt to assess the impact of the intervention on migration. Migration in this section is narrowly defined as people leaving their households, regardless of the location of destination. A second definition is provided that refers to individuals leaving the household to relocate outside the Northern region. This latter definition contains both elements of distance and long-term movement that we normally associate with migration. We then look at the reported reasons for migrating.

The project does not have any impact on migration so narrowly defined (Table 63). It increases the number of people leaving the household but only in the last survey round. It has an impact on individuals moving away from the household in order to care for other relatives or for 'other' reasons. The project has no impact at all on migration outside the Northern region of the country.

Table 63 Impact of MV on migration

	Baseline CV	Baseline diff. MV	DD impact 2013	DD impact 2014	DD impact 2015	DD impact 2016	DD average impact
Migrants per household	0.32	0.08	-0.04	0.03	-0.07	0.35**	0.09
		(0.311)	(0.695)	(0.779)	(0.473)	(0.008)	(0.351)
Migrants outside the	0.20	0.05	0.00	0.04	-0.06	0.02	0.00
Northern region		(0.373)	0.995	(0.638)	(0.369)	(0.850)	(0.982)
Work migrants per	0.17	0.06	-0.04	-0.08	-0.08	-0.06	-0.06
household		(0.195)	(0.523)	(0.136)	(0.161	(0.258)	(0.171)
Study migrants per	0.06	0.01	0.02	-0.03	0.02	0.04	0.01
household		(0.796)	(0.431)	(0.500)	(0.570)	(0.309)	(0.689)
Care migrants per	0.08	0.03	0.05	0.06	0.01	0.19**	0.08*
household		(0.392)	(0.266)	(0.345)	(0.843)	(0.010)	(0.074)
Other migrants	0.02	-0.01	0.01	0.09**	-0.020	0.18**	0.06**
		(0.174)	(0.757)	(0.021)	(0.433)	(0.002)	(0.004)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

We operate a further distinction between 'temporary' and 'permanent' migrants. The distinction is based on a survey question on whether the person who moved intends to stay away from the original household for a year or more. We coded a positive answer to this question as a permanent move away from the original family. Interestingly, after performing this split, we find a positive impact of MV on permanent migration and a negative impact on temporary migration (Table 64).

This result may simply be a statistical accident resulting from slicing the sample, but if we are to believe these data, they suggest MV is increasing the number of people leaving their households for a year or more, including to move to areas outside the Northern region. As for the reasons, care, work and study are, in order, the most important. On the other hand, MV is decreasing temporary migration, particularly for working reasons, suggesting the project is reducing seasonal migration of labourers.

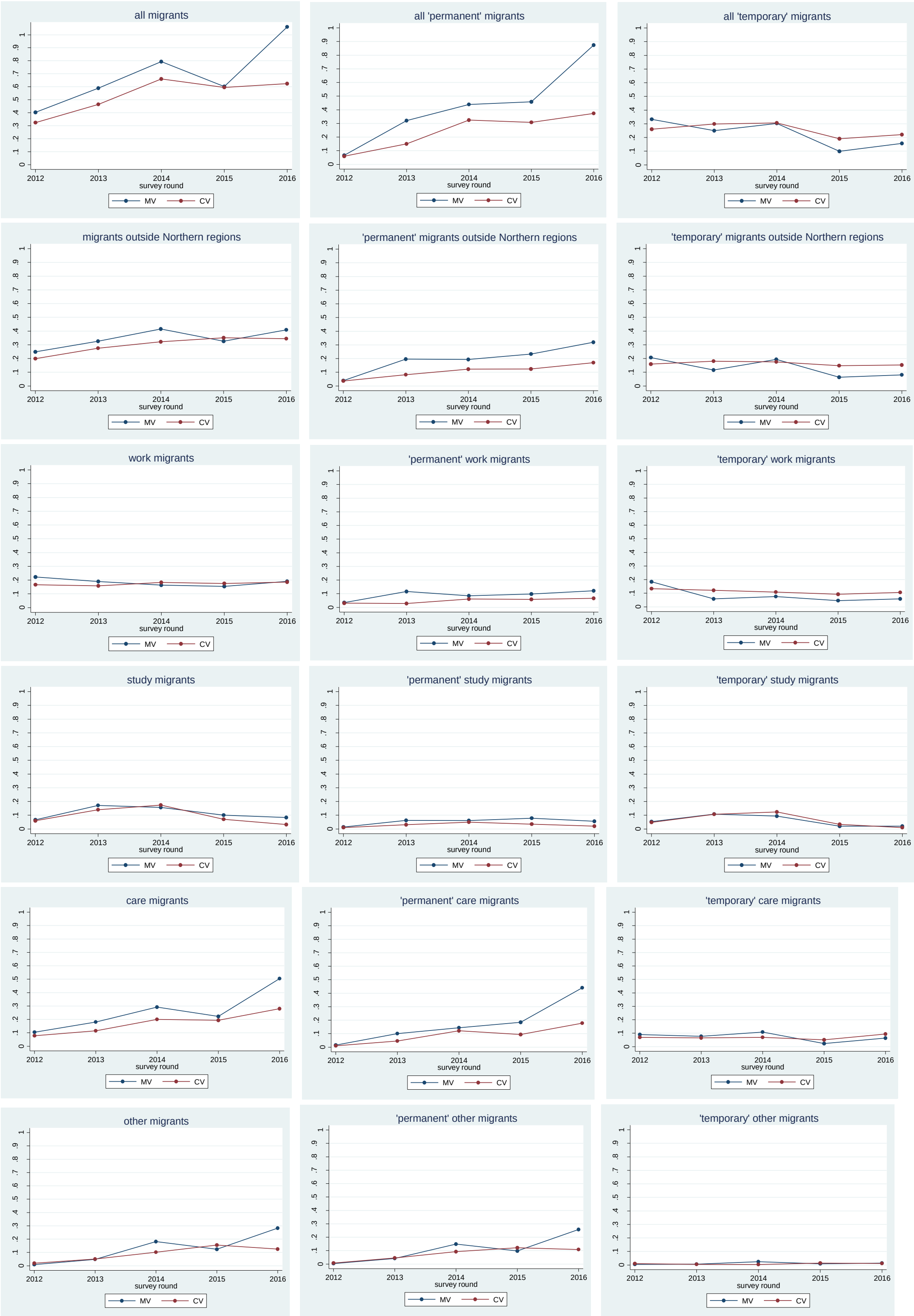
Table 64 Impact of MV on temporary and permanent migrants

MDG	Baseline diff. MV	DD average impact all migrants	Baseline diff. MV	DD average impact 'permanent' migrants	Baseline diff. MV	DD average impact 'temporary' migrants
Migrants per household	0.08	0.09	0.01	0.21***	0.07	-0.12*
	(0.311)	(0.351)	(0.717)	(0.000)	(0.290)	(0.062)
Migrants outside the	0.05	0.00	0.00	0.10**	0.05	-0.08**
Northern region	(0.373)	(0.982)	(0.885)	(0.009)	(0.322)	(0.047)
Work migrants per	0.06	-0.06	0.00	0.04*	0.05	-0.09**
household	(0.195)	(0.171)	(0.793)	(0.080)	(0.152)	(0.006)
Study migrants per	0.01	0.01	0.00	0.03*	0.01	-0.02
household	(0.796)	(0.689)	(0.751)	(0.065)	(0.818)	(0.442)
Care migrants per	0.03	0.08*	0.00	0.10***	0.02	-0.02
household	(0.392)	(0.074)	(0.592)	(0.000)	(0.485)	(0.575)
Other migrants	-0.01	0.06**	0.00	0.05**	-0.01	0.01
	(0.174)	(0.004)	(0.487)	(0.016)	(0.321)	(0.108)

Note: Coefficients are DD estimates obtained using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. P-values in parentheses based on cluster standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance

One limitation of this analysis lies in the definition of migration used. In principle we would like to analyse the proportion of individuals moving across MV and CV areas and outside. To do this we need to code the information on the location of destination. This information is available but identifying and coding villages of destination as MV, CV and other is a demanding task that we postpone for future research.

Figure 69 Migration patterns in MV and CV areas



13. Conclusions

In this section we briefly summarise some of the most obvious conclusions emerging from the quantitative analysis. Despite the initial scepticism advanced by many sides, we were able to build a valid counterfactual of comparable control villages and to collect data from a large panel of households. Though sensitivity of estimates to covariate shocks cannot be ruled out, we were able to collect data from control villages that were similar to project villages in baseline characteristics and in trends. The attrition among the sample of panel households was minimal and the quality of the data collected is somewhat inferior but similar to the quality of similar data collected in Ghana by other institutions such as GSS.

First, participation in most project activities was high. MV activities clearly reached a large proportion of the population. The activities promoted by the intervention changed over time as some activities were discontinued while others were started from scratch along the way, but participation was always significantly higher than participation observed in similar activities in control areas. Participants in project activity did not display any particular characteristics such as being poorer or less educated than average. The project appeared to be followed by large numbers of households without being targeted to any specific group.

Second, the project did not produce the expected positive and large results in terms of the MDGs. While some MDG indicators were favourably affected, others were not. In addition, the impacts were relatively small and achieved statistical significance only for the following outcomes: attendance rates in primary education, rates of birth attendance by skilled professionals, prevalence of contraception, rate of children sleeping under insecticide-treated bednets and access to improved sanitation facilities. Once the improvements are aggregated in single summary index such as the Oxford MPI, MV appears to be overall successful, though the practical significance of the improvement in the aggregate index measure is difficult to interpret.

Third, the project did not seem to generate the expected synergistic effects. Simultaneous investments in several sectors do not appear to have produced a dramatic change in living conditions. This is suggested by the small impacts produced on the final outcomes despite the wide reach of project activities and by the total absence of impact on ultimate indicators of success such as monetary poverty and child mortality. In the only case in which we were able to assess the presence of synergies (estimation of the impact of MV on agricultural production), we were able to find only a limited ‘unexplained’ impact that could be attributed to improvements occurring outside agriculture.

Fourth, the project did not reduce monetary poverty but it improved household incomes. We explain this apparent paradox as the combined result of two factors. First, incomes are underestimated by the survey, and the reported increase in agricultural income, which was affected by the intervention, overstates the actual change in overall household income. Second, people saved rather than spent the income gains generated by the project. This is consistent with one version of the permanent income hypothesis of consumption whereby individuals do not spend income gains that do not represent permanent changes in income. In other words, households perceived the income gains made, perhaps correctly, as temporary gains and decided to save them for precautionary reasons.

Fifth, the MV produced a considerable impact on a number of non-MDG welfare indicators. In particular, it significantly improved agricultural incomes and savings, self-reported food security and stunting rates, and improved somewhat the incidence of malaria and health behaviours like hand washing, child vaccinations and breastfeeding. It also appeared to reduce temporary seasonal migration while increasing permanent migration outside the study area. Given the lack of data on intermediate outcomes, we were mostly unable to explain the operation mechanisms of these changes. Impacts in agriculture appear to be driven mostly by the larger use of chemical fertiliser, pesticides and seeds over larger land plots rather than by increases in land productivity. Failure to translate the health interventions into improvements in final outcomes may have to do with the difficulty to change people’s knowledge, attitudes and behaviours.

Sixth, spill-over effects to neighbouring areas did not materialise. The small observed impact of the intervention is not dampened by the occurrence of an impact in nearby control villages, as the latter were scarcely positively affected by the interventions. In addition, the impact on neighbouring villages did not always have the expected positive sign, suggesting that in some districts and for some interventions the spill-over effects could have been negative. Negative spill-over effects are plausible if the district resources deployed in MV areas are taken away from neighbouring areas. Similarly, we could not find significant heterogeneous effects across genders or districts.

The project did not perform better or worse in one district compared with the other, while girls, women and female-headed households did not benefit from the intervention more or less than males.

Appendix A: Matching methods

Imbens and Rubin (2015) recommend the separation of the design stage from the analysis stage in conducting observational studies. The goal of the 'design' stage is to select a propensity score and a sample of observations that maximises the statistical balance of the distribution of the covariates. In the design stage the outcomes are completely ignored in order not to bias the construction of the propensity score. The goal of the analysis stage is to estimate project effects using the propensity score estimated in the design stage. We briefly describe the various steps followed in the design and analysis stage.[17]

At the design stage we estimate the propensity score using a logistic regression model. Imbens and Rubin (2015) propose an algorithm for the estimation of the propensity score that aims at achieving statistical balance of the covariates and does not try to 'explain' participation through a behavioural model. After estimating the propensity scores, we assess their validity, the balancing of the covariates and the overlap in the distribution of the covariates. We then trim the data and re-estimate the propensity scores on the trimmed sample. We then use the re-estimated propensity score to build groups of similar project and comparison observations and we estimate project effects within these groups using linear regression. The regressions include additional sets of covariates that are relevant determinants of the outcome considered and that are not being affected by the project. The population-level project effect is obtained as a weighted average of the project effect calculated within groups.

In this appendix we also assess the validity of the unconfoundedness assumption by running two placebo tests: the impact of the intervention on outcomes that are known not to be affected by the intervention and the impact of the intervention between two areas of the control group in which the intervention was not implemented. We then assess the robustness analysis assessing the sensitivity of the results to the choice of covariates.

Our first goal is to estimate a propensity score that balances the covariates in the project and the comparison groups. In other words, we want to estimate a propensity score such that, within sub-samples with similar values of the estimated propensity score, the covariates of the project and control group are similar. To do so we start with a model based on substantive knowledge and we refine the model based on its ability to achieve balance in the covariates within strata defined using an estimated propensity score. We are not building a causal model of the propensity score explaining the selection of the MV areas, partly because the choice of the MV areas was not based on a pre-specified set of variables. The area comprising the MV areas was chosen using poverty maps and stakeholder consultations without relying on specific indicators that can be reconstructed in a probabilistic selection model. Our aim in estimating a propensity score is to achieve adequate balance between the covariate distribution of the project and the control groups.

Estimating the propensity score

We estimate the propensity score using logistic regression and a set of covariates selected in the following way. We identify a number of basic covariates X_b that we include in the model regardless of their explanatory power because we believe they are strong determinants of all outcomes considered. We then include additional covariates X_a from the full pool of potential covariates X that are not affected by the intervention, based on their statistical significance. Finally, we include square terms and interactions of the basic X_b and additional covariate X_a.

The covariates included in the model consist of factors that are not affected by the programme. This is ensured by design because we only use baseline values of the covariates before the project started. The full list of covariates is composed of household characteristics that are likely to be strongly correlated with many outcomes affected by the programme. Five covariates (basic covariates) are included a priori: household size, age of head of household, education of head of household, size of cultivated land and value of total household wealth (livestock plus durable assets and productive assets). The pool of potential covariates includes 24 household variables: household isolation (not having relevant ties with other households), polygamous households, female-headed households, having at least one household member migrated for work, having a member sending

[17] The full procedure for estimating the propensity score and assessing covariate balance is implemented by running sequentially the following stata dofiles: pscore.do, blocks.do, balance.do, overlap.do, trim.do, pscore2.do, balance.do and overlap.do

remittances, not having access to protected water, distance to nearest source of drinking water, not having access to a protected toilet, running a community service business, running a trade business, running a small food business, running any other business, being affected by a drought in the past three years, walls made of mud, floor made of earth, roof made of metal, farmer household, main crop is maize, main crop is millet, main crop is rice, main crop is groundnut, number of months food-insecure, having bank savings and being member of susu. All the potential covariates are binary with the exception of distance to nearest source of drinking water and number of months food-insecure. The potential covariates are included in the model stepwise provided they achieve a level of statistical significance equivalent to a P-value below 15%. The results of this first step are presented below. A total of 19 covariates are included in the model, of which 5 are basic and 14 are additional (see Table A1)

Table A1 Logistic regression of the propensity score (basic and additional variables)

Variable	Coefficient	s.e.	T-stat.	P-value
Household size	-0.007	0.014	-0.500	0.617
Age head	0.004	0.003	1.150	0.250
Education head	0.000	0.017	0.020	0.980
Cultivated land	0.053	0.022	2.450	0.014
Wealth	0.000	0.000	0.730	0.468
Months food insecure	-0.123	0.030	-4.060	0.000
Remittances	0.927	0.281	3.300	0.001
Isolated household	0.345	0.119	2.910	0.004
Flood shocks	0.648	0.146	4.440	0.000
Millet farm	0.417	0.113	3.680	0.000
Rice farm	-0.477	0.109	-4.390	0.000
Drought shock	-0.510	0.146	-3.480	0.000
Groundnut farm	0.212	0.101	2.110	0.035
Farmer household	-0.392	0.168	-2.340	0.019
Distance to water	0.003	0.001	1.910	0.056
Bank access	0.383	0.149	2.580	0.010
Metal roof	-0.265	0.106	-2.500	0.012
Maize farm	0.265	0.122	2.170	0.030
Working migrant	0.238	0.155	1.540	0.125
Constant	-0.990	0.299	-3.310	0.001
Observations	2,172			
Pseudo R2	0.052			

We then expand the model to include square terms and interactions. Since most covariates are binary variables we have only seven square terms. In order to avoid multicollinearity we do not include the full range of interactions and we restrict interactions to interactions with the Builsa region dummy variable. Squares and interactions are added stepwise to the previous model specification using a cut-off of significance level equivalent to a P-value of 5%. This procedure leads to the inclusion of five square terms and nine interactions terms to the previous model. The final model therefore includes a total of 33 estimated coefficients (see Table A2).

Table A2 Logistic regression of the propensity score augmented by squares and interaction terms

Variable	Coefficient	s.e.	T-stat.	P-value
Household size	0.107	0.041	2.610	0.009
Age head	0.003	0.003	0.950	0.343
Education head	0.262	0.059	4.430	0.000
Cultivated land	0.003	0.026	0.130	0.893
Wealth	0.000	0.000	2.010	0.044
Months food insecure	0.142	0.082	1.720	0.085
Remittances	0.821	0.290	2.830	0.005
Isolated household	-0.207	0.183	-1.130	0.257
Flood shocks	-2.178	0.549	-3.970	0.000
Millet farm	0.458	0.143	3.200	0.001
Rice farm	0.180	0.180	1.000	0.317

Drought shock	-1.611	0.210	-7.650	0.000
Groundnut farm	0.241	0.109	2.210	0.027
Farmer household	-0.459	0.180	-2.550	0.011
Distance to water	0.000	0.002	0.150	0.882
Bank access	0.442	0.158	2.800	0.005
Metal roof	0.046	0.146	0.310	0.754
Maize farm	0.319	0.144	2.210	0.027
Working migrant	0.299	0.162	1.850	0.065
Drought shock X Builsa	2.007	0.275	7.300	0.000
Rice farm X Builsa	-1.349	0.240	-5.630	0.000
Flood shock square	2.548	0.492	5.180	0.000
Food insecure X Builsa	-0.237	0.067	-3.550	0.000
Isolated household X Builsa	1.204	0.254	4.730	0.000
Education head X Builsa	-0.144	0.039	-3.700	0.000
Education head square	-0.019	0.005	-3.840	0.000
Household size square	-0.006	0.002	-3.000	0.003
Cultivated land X Builsa	0.201	0.050	4.020	0.000
Metal roof X Builsa	-0.712	0.226	-3.150	0.002
Millet farm X Builsa	-0.781	0.251	-3.120	0.002
Food insecure square	-0.033	0.014	-2.310	0.021
Distance to water X Builsa	0.006	0.003	2.050	0.041
Wealth square	0.000	0.000	-1.870	0.061
Constant	-0.683	0.371	-1.840	0.066
Observations	2,172			
Pseudo R2	0.136			

Building strata

The second step consists of using the estimated propensity score to construct strata in such a way that within each stratum the variation in the estimated propensity score is small. To do so we adopt an iterative procedure. But first we make two adjustments to the data. First, we linearise the propensity score (or log odds ratio, le): *le=ln(e/(1-e),* because the linearised propensity score is more likely to have a distribution that is well approximated by a normal distribution. Second, we use only observations overlapping in the propensity score. That is, we drop control units with a propensity score lower that the smallest estimated propensity score in the project group, and we drop project units with a propensity score larger than the largest value of the propensity score in the control group.

A stratum is considered adequate if a T-test between project and control group in the strata is below a given threshold. The low value of the T-test suggests that, within the stratum, conditional on the propensity score, the covariates are independent on the treatment. We select a threshold value for the T-test of 1.69. If the T-test is larger than 1.69 the group is split into two other groups provided that each project and control sub-group contains a pre-specified minimum number of observations. The split is made calculating the median of the estimated propensity score in the stratum and then creating four groups of project and control observations below and above the median. In our application we use a minimum of 20 observations to make sure that estimation of project effects within a stratum can be conducted with a sufficiently large number of additional covariates. The process is repeated iteratively until either the T-test is lower than 1.69 or there are fewer than 20 observations in the control or project group in a given stratum.

Table A3 Optimum sub-classification based on the propensity score

Stratum	Obs.	Project	Control	P-score P	P-score C	T-test	P-value
1	529	74	455	-1.891	-1.948	1.32	0.189
2	264	54	210	-1.230	-1.230	0.21	0.834
3	265	68	197	-1.012	-1.025	1.12	0.263
4	529	183	346	-0.581	-0.556	1.46	0.144
5	264	141	123	0.010	0.010	0.03	0.978
6	132	85	47	0.506	0.501	0.23	0.815
7	132	98	34	1.194	1.111	1.18	0.241

The routine produced seven strata that are presented in Table A3 ordered by the value of the linearised propensity scores. The smallest stratum contains 132 observations and the smallest group contains 34 observations. Values of average propensity score are fairly similar within strata and all the P-values of statistical tests of the difference across project and control groups are larger than 14%.

Assessing covariates' balance

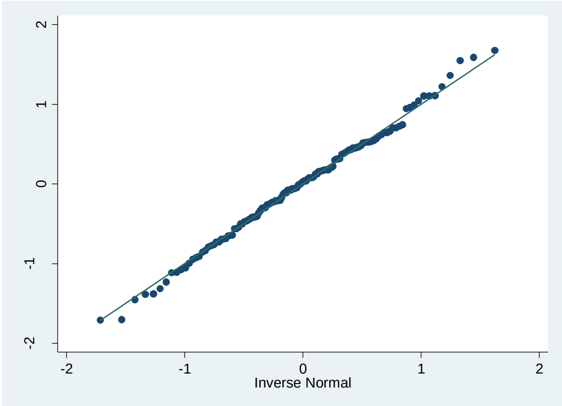
We then proceed to test the hypothesis that covariates are independent of the treatment conditional on the propensity score: W _||_ X|e. To do so we conduct three different tests. The first test assesses the global balance of each covariate across strata. In this test each covariate in turn is estimated as an average pseudo-effect of the project. The test reports Z-values of the null hypothesis that the pseudo-effect is equal to zero. The Z-values should follow random draws from a normal distribution. If there are several large Z-values the null hypothesis is rejected. In the second test we assess the balance for each covariate within all strata. We test for each covariate that its project mean is equal to the control mean in each stratum. To do this we estimate linear regressions of the outcomes on the strata dummies and the interactions' strata-project dummies. We report an F-test of the slope dummies and the corresponding P-value. In the third test, we assess balance within strata for each covariate. Since there are 19 covariates and 7 strata, this delivers 133 different tests. If covariates are well balanced we would expect Z-values to have smaller values than those expected from a normal distribution. Instead of reporting each single test, we plot the Z-values using a Q-Q plot. If the variables are well balanced we would expect the plot to lie above the 45 line.

Table A4 Measures of covariance balance

Covariate	Unadjusted T-test	Z-value across strata	F-value within all strata	P-value within all strata
Household size	-0.83	-0.52	1.42	0.19
Age of head	-1.50	-0.05	0.80	0.59
Education of head	-0.98	-0.30	1.11	0.35
Cultivated land	-2.63	-0.48	1.03	0.41
Wealth	-1.78	-0.21	0.37	0.92
Food security	3.61	0.28	0.20	0.99
Remittances	-4.19	0.03	2.44	0.17
Isolated household	-3.56	0.00	0.66	0.71
Flood shocks	-3.25	-0.19	1.02	0.48
Millet farm	-2.82	0.14	0.27	0.97
Rice farm	3.92	0.42	0.50	0.83
Drought shocks	3.02	-0.33	3.24	0.00
Groundnut farm	-1.95	0.43	0.76	0.62
Farmer household	2.52	0.03	0.27	0.97
Distance to drinking water	-1.49	-0.04	1.35	0.22
Bank access	-3.13	-0.32	0.44	0.88
Metal roof	0.50	-0.20	0.31	0.95
Maize farm	-2.08	-0.31	0.67	0.70
Migrants	-3.59	-0.31	0.93	0.48

The results of theses tests are reported in Table A4. The first column shows the standard t-test of differences in means applied to the unadjusted data. As expected from non-experimental data, several of the differences are statistically significant. The Z-values of the equality tests across strata are very low and lower than one would expect from random draws from a normal distribution of Z-values. The last two columns report F-tests and P-values of slope dummies of project-strata interactions for each covariate. The tests for each single covariate and stratum are the reported in a Q-Q plot (see Figure A1). Most observations lie above or on the 45 line suggesting that the results are very similar to those obtained from random draws of Z-values from a normal distribution. Note that, for this latter test, the distribution of the covariates would be more balanced if we had included squared terms and interaction terms.

Figure A1 Quantile-quantile plot of Z-scores



Assessing the overlap in the distribution of the covariates

Causal effects from non-experimental data can be estimated using different models, including simple linear regression, propensity score matching and sub-classification. If the distribution of the covariates in the project and control groups is similar, the estimated effects are less likely to be sensitive to the choice of the estimation model. In this session we assess the degree of covariate balance, and the extent of overlap in the distribution of covariates.

We assess the balance in the distribution of covariates using three summary measures. The first is the normalised difference. This is the difference in the means in the two groups divided by the square root of the average of the two within group variances (equation 14.1 in Imbens and Rubin, 2015). This measure is preferable to standard t-test statistics because our interest is in the size of the difference between covariates regardless of sample size. Imbens and Rubin make a nice example in this regard. If you preserve the same average distance and distribution while doubling the sample size, the t-statistic will increase while the standardised difference remains unchanged. The second measure is the difference in the logarithms of the standard deviations of the covariate in the project and the control group (equation 14.4 of the book). This measure compares the dispersion of the two distributions for each covariate. We use the difference in the logarithms rather than the difference in the standard deviations because it is more likely to be normally distributed. The third measure is the proportion of control and treated units with covariate values outside the 2.5% and 97.5% quantiles of the distribution of the covariate values in the treated and control units (equations 14.6 and 14.7 of the book). This approach investigates the proportion of project (or control) units with covariate values in the tails of the distribution of the covariate values. Since many of the covariates used are binary, this latter measure is not very informative and we report the results only for continuous covariates.

It should be noted that any imbalance in the covariate distribution, whether in expectations, in dispersion or in the shape of the distribution, leads to a difference in the distribution of the propensity scores. Therefore we report the three measurements above for the linearised propensity score in the first place as summary measures of the overall balance of the covariate distribution.

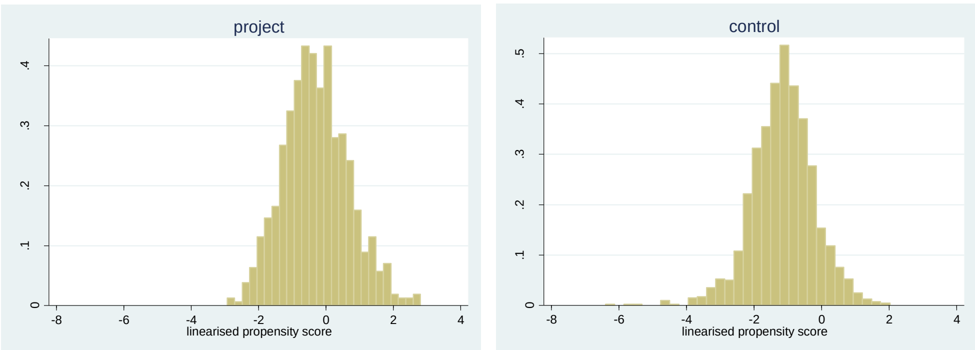
Table A5 Measures of covariate balance in the distributions

Covariate	Normalised difference	Difference in standard deviations	Project proportion outside 95% of distribution	Control proportion outside 95% of distribution
Linearised propensity score	0.910	0.036	0.153	0.138
Household size	0.050	-0.096	0.031	0.057
Age of head	0.066	0.013	0.052	0.034
Education of head	0.032	-0.040	0.001	0.032
Cultivated land	0.141	0.060	0.030	0.018
Wealth	0.076	-0.064	0.066	0.044
Food security	-0.235	-0.150	0.001	0.031

Covariate	Normalised difference	Difference in standard deviations	Project proportion outside 95% of distribution	Control proportion outside 95% of distribution
Remittances	0.198	0.509	0.059	0
Isolated household	0.187	0.140	0	0
Flood shocks	0.146	0.046	0	0
Millet farm	0.118	-0.042	0	0
Rice farm	-0.206	-0.077	0	0
Drought shocks	-0.137	0.134	0	0
Groundnut farm	0.102	-0.008	0	0
Farmer household	-0.115	0.160	0	0
Distance to drinking water	0.082	0.273	0.027	0.012
Bank access	0.144	0.159	0	0
Metal roof	-0.037	-0.009	0	0
Maize farm	0.128	-0.069	0	0
Migrants	0.181	0.193	0	0

It is also useful to construct histograms of the distribution of the covariates in the project and control groups separately to spot differences in the shape and the distribution that are not immediately captured by differences in the means and the variances. Figure A2 shows histograms for the linearised propensity scores in the project and control areas, respectively.

Figure A2 Distributions of linearised propensity scores in project and control group



Improving balance by trimming

Balance in the covariates can be further improved by dropping from the sample those observations whose covariates are very different between project and control groups. Trimming the sample improves our estimates of the causal effects and changes the values of the estimated outcomes. This however, is achieved at the expense of external validity. Inferences based on a smaller sample have lower general validity and are more difficult to extrapolate to other contexts. Trimming is conducted to improve internal validity. To illustrate the problem, suppose there are observations whose propensity scores are exactly 0 or 1. For these observations there are no valid comparators in the project or control group. A similar reasoning applies to observations whose values of the propensity score are very close to 0 or 1. Control comparators for project observations with a probability of being in the sample larger than 99% are hard to find by definition, because, among 100 observations with a propensity score of 0.99, only 1 will be from the control group. In this section we first assess whether trimming is needed and then proceed to trim the sample to improve the balance of the covariates. In order to identify observations that are too close to 0 and 1 we identify a threshold relying on the asymptotic sampling variance of estimators for average treatment effects. All units whose propensity score is in the intervals [0,a] and [1-a,1] are discarded, while causal effects are estimated only for the observations in the interval [a,1-a]. First, we assess whether trimming is needed by checking the inequality (equation 16.9 of Imbens and Rubin, 2015). Second, if the inequality does not hold, we find the threshold value of a, using equation 16.10 of Imbens and Rubin.

The sample data fail the inequality test. The difference between the right hand side and the left hand side of equation 16.9 is negative (=585.68) suggesting that trimming is needed. We then identify the value of *a* defining the interval [0.1001, 0.8998]. This interval drops 170 observations, corresponding to 7.8% of the original sample, of which 18 are from the project group and 152 from the control group. The trimmed sample therefore comprises 2,002 observations. We repeat the test above to assess the need for trimming and the inequality is now respected as it delivers a positive value (0.04).

Re-estimation of the propensity score

After trimming the sample we re-estimate the propensity score using the same procedure outlined above. With the new propensity score we build new sub-classes for the estimation of project effects. We also repeat many of the same tests conducted above, which, after trimming and re-estimation of the propensity score, improve considerably. A total of 17 covariates are included in the model, of which 5 are basic and 12 are additional (see Table A6).

Table A6 Logistic regression of the propensity score (basic and additional variables)

Variable	Coefficient	s.e.	T-stat.	P-value
Household size	-0.001	0.014	-0.090	0.928
Age of head	0.002	0.003	0.570	0.571
Education of head	0.017	0.019	0.900	0.368
Cultivated land	0.054	0.022	2.480	0.013
Value of wealth	0.000	0.000	0.610	0.540
Remittances	1.025	0.258	3.980	0.000
Millet farm	0.389	0.112	3.480	0.000
Rice farm	-0.399	0.111	-3.600	0.000
Drought shock	-0.531	0.148	-3.580	0.000
Flood shock	0.578	0.147	3.940	0.000
Isolated household	0.311	0.119	2.610	0.009
Months food insecure	-0.089	0.032	-2.760	0.006
Farmer	-0.393	0.170	-2.320	0.021
Bank access	0.325	0.150	2.170	0.030
Metal roof	-0.187	0.109	-1.720	0.086
Distance to water	0.002	0.001	1.640	0.101
Groundnut farm	0.167	0.102	1.630	0.102
Constant	-0.685	0.293	-2.340	0.019
Observations	2,002			
Pseudo R2	0.042			

We then expand the model to include square terms and interactions. Since most covariates are binary variables we only have seven square terms. In order to avoid multicollinearity we do not include the full range of interactions but only interactions with the regional dummy for the Builsa region. Squares and interactions are introduced stepwise to the previous model specification using a cut-off of significance level equivalent to a P-value of 5%. This procedure leads to the inclusion of five square terms and ten interactions to the previous model. The final model therefore includes a total of 32 estimated coefficients (see Table A7).

Table A7 Logistic regression of the propensity score augmented by squares and interaction terms

Variable	Coefficient	s.e.	T-stat.	P-value
Household size	0.082	0.043	1.890	0.059
Age head	0.002	0.004	0.620	0.532
Education head	0.285	0.064	4.470	0.000
Cultivated land	-0.010	0.028	-0.370	0.709
Wealth value	0.000	0.000	3.130	0.002
Remittances	1.019	0.267	3.820	0.000
Millet farm	0.291	0.131	2.230	0.026
Rice farm	0.180	0.182	0.990	0.322
Drought shock	-1.690	0.225	-7.510	0.000
Flood shock	-2.414	0.559	-4.310	0.000
Isolated household	-0.180	0.186	-0.970	0.332
Months food insecure	0.102	0.087	1.170	0.242
Farmer	-0.495	0.182	-2.720	0.006

Variable	Coefficient	s.e.	T-stat.	P-value
Bank access	0.447	0.160	2.790	0.005
Metal roof	-0.002	0.148	-0.020	0.988
Distance to water	0.000	0.002	-0.030	0.973
Groundnut farm	0.242	0.110	2.190	0.028
Isolated household X Builsa	1.212	0.259	4.680	0.000
Flood shock square	2.763	0.503	5.500	0.000
Rice farm X Builsa	-1.269	0.246	-5.160	0.000
Drought shock X Builsa	2.155	0.313	6.890	0.000
Builsa	-1.255	0.345	-3.640	0.000
Cultivated land X Builsa	0.257	0.055	4.630	0.000
Education head square	-0.023	0.006	-4.080	0.000
Age head squared	-0.133	0.044	-3.000	0.003
Wealth value square	0.000	0.000	-2.550	0.011
Wealth value X Builsa	0.000	0.000	-2.300	0.021
Months food insecure square	-0.033	0.016	-2.060	0.039
Distance to water X Builsa	0.007	0.003	2.260	0.024
Household size square	-0.005	0.002	-2.120	0.034
Metal roof X Builsa	-0.542	0.236	-2.300	0.022
Months food insecure X Builsa	-0.161	0.073	-2.220	0.026
Constant	0.113	0.393	0.290	0.773
Observations	2,002			
Pseudo R2	0.112			

Building strata II

We then proceed to build new sub-classes of the sample observations using the same procedure employed above. The routine produced eight strata (see Table A8). Table A8 reports the number of observations in each sub-class and by project assignment, the average propensity score in the project and control groups and T-test and P-values of the equality of the propensity scores between project and control group within each sub-class. The smallest stratum contains 125 observations and the smallest group contains 29 observations. Average values of propensity score in the project and control groups are nearly identical up to two digits in all sub-classes. T-statistics are low and differences are never statistically significant.

Table A8 Subclasses bases on the re-estimated propensity score

Stratum	Obs.	Project	Control	P-score P	P-score C	T-test	P-value
1	501	80	421	0.16	0.16	-0.96	0.336
2	250	51	199	0.23	0.23	-1.05	0.295
3	250	72	178	0.28	0.28	-0.71	0.476
4	251	82	169	0.34	0.33	-0.96	0.338
5	250	104	146	0.41	0.41	0.39	0.698
6	250	127	123	0.51	0.51	-0.29	0.772
7	125	81	44	0.63	0.62	-1.32	0.191
8	125	96	29	0.76	0.75	-1.22	0.225

Assessing covariates' balance II

We reassess balance of the covariates after re-estimating the propensity score as before.

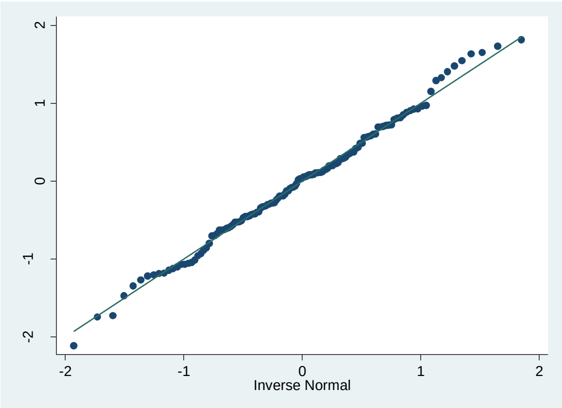
Table A9 Measures of covariance balance

Covariate	Unadjusted T-test	Z-value across strata	F-value within all strata	P-value within all strata
Household size	-0.62	-0.13	0.71	0.685
Age of head	-1.17	0.25	1.42	0.182
Education of head	-1.49	-0.24	1.14	0.335
Cultivated land	-2.29	-0.31	0.47	0.878
Wealth	-1.49	0.01	0.69	0.701
Remittances	-3.91	-0.03	2.19	0.025
Millet farm	-3.22	-0.10	0.75	0.651
Rice farm	3.01	0.09	0.43	0.904
Drought shocks	3.37	-0.07	3.75	0.000

Covariate	Unadjusted T-test	Z-value across strata	F-value within all strata	P-value within all strata
Flood shock	-2.67	-0.03	1.40	0.199
Isolated household	-3.24	-0.03	0.44	0.899
Months food-insecure	2.79	0.12	0.44	0.897
Farmer household	2.58	0.27	1.07	0.380
Bank access	-2.84	-0.06	1.02	0.421
Metal roof	0.53	-0.09	0.52	0.840
Distance to drinking water	-1.13	0.00	0.59	0.790
Groundnut farm	-1.88	0.46	1.26	0.260

As before, we plot all the Z-scores using a QQ plot to check whether they conform to a normal distribution.

Figure A2 Quantile-quantile plot of Z-scores



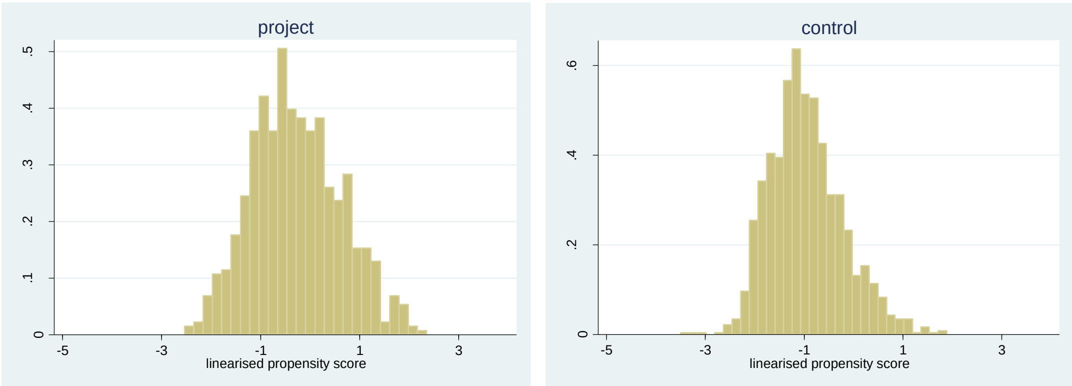
Assessing the overlap in the distribution of the covariates II

Finally, as before, we reassess the balance and the overlap in the distribution of the covariates and we plot the distribution of the linearised propensity scores using histograms.

Table A10 Measures of covariate balance in the distributions

Covariate	Normalised difference	Difference in standard deviations	Project proportion outside 95% of distribution	Control proportion outside 95% of distribution
Linearised propensity score	0.821	0.190	0.163	0.080
Household size	0.029	-0.057	0.033	0.053
Age of head	0.055	-0.000	0.052	0.036
Education of head	0.070	0.031	0.017	0.024
Cultivated land	0.107	0.021	0.027	0.022
Wealth	0.070	-0.047	0.051	0.058
Remittances	0.172	0.432	0.056	0
Millet farm	0.153	-0.052	0	0
Rice farm	-0.143	-0.059	0	0
Drought shocks	-0.155	0.142	0	0
Flood shock	0.124	0.044	0	0
Isolated household	0.145	0.108	0	0
Months food insecure	-0.132	-0.035	0.004	0.009
Farmer household	-0.118	0.165	0	0
Bank access	0.130	0.142	0	0
Metal roof	-0.025	-0.007	0	0
Distance to drinking water	0.051	0.227	0.026	0.018
Groundnut farm	0.088	-0.007	0	0

Figure A3 Distributions of the linearised propensity scores in the project and control group



Analysis

Our method of choice for the analysis is sub-classification, also referred to as blocking or stratification. The assumption of this method is that, within sub-groups with the same value of the propensity score, the distribution of the covariates is identical in the project and control groups. In the case of some indicators that cannot be estimated by a simple regression or a comparison of means such as the poverty gap and mortality rates, we employ weighting estimators.

Sub-classification

In this method, after trimming the data in the way described above, we subdivide the data based on the values of the linearised estimated propensity score on the trimmed data.

After constructing the strata as in Table A8, we analyse the data as if they were from a stratified randomised trial, assuming that units within the same stratum have the same propensity score. The simplest way to do so consists of estimating the average impact of the project as the difference between the average outcome in the project group and the average outcome in the control group in each stratum *j*:

$$d_j = \overline{Y_{1J}} - \overline{Y_{0J}}$$

We then average the project effects over all strata by weighting each stratum effect by the size of the stratum (*Nj/N*). This produces the average treatment effect of the intervention (ATE):

$$d = \sum_{j=1}^{J} \frac{N_j}{N} d_j$$

If we are interested in the average treatment effect on the treated (ATT) we use the proportion of the project units in each stratum rather than the using the size of the stratum:

$$d = \sum_{j=1}^{J} \frac{N_{1j}}{N_1} d_j$$

The estimates of the project effects can be improved by estimating project effects within strata using regression analysis:

$$y_i = \alpha_j + d_j P_i + \beta_j X_i + e_i$$

The inclusion of covariates in the regression has two effects. First, it increases the precision of the estimates. Second, it reduces bias. The difference in the outcome means between project and control group within each stratum is likely to be biased because the propensity scores are only approximately similar within the blocks. If the propensity score balanced the covariates perfectly, then the inclusion of the covariate in the regression above would be irrelevant to the estimation of the project effects (*d*'s) and the estimated beta would be zero. But normally the inclusion of the covariate will reduce bias in the estimate of d resulting by existing imbalances between the covariates that were not fully resolved by stratifying the observations using the propensity score. To estimate project effects within strata we employ a cross-sectional model, fixed effect models and a lagged dependent variable model depending on the availability of panel data. Once the project effects are estimated within each stratum, they are averaged across strata to produce average treatment effects or average treatment effects on the treated. The standard errors of the project effects are then calculated using 500 bootstrapped replications.

Weighting

In this approach, the inverse of the propensity score is used to weight observations in order to remove the bias resulting from differences in the covariates between the project and the control group. Weights can be used in the estimation of mean differences or in weighted regression analysis. The use of weighting by the propensity score in regression analysis of project effect is considered 'double robust' (Imbens & Wooldridge, 2009) because the weighted regression estimator is consistent as long as the specification of the propensity score or the specification of the regression function are correct. The weights for the estimation of average treatment effects are:

$$w_{ATE} = \frac{P}{e} + \frac{1-P}{1-e}$$

while weights for the estimation of the treatment effect on the treated are:

$$w_{ATT} = P + (1-P)\frac{e}{1-e}$$

In some cases, estimates based on weighting and sub-classification are very similar, for example when sub-classification uses many strata and the dispersion of the propensity score within each stratum is limited, or when the overall variation in the propensity score is small or when there are few extreme values of the propensity score. In general, however, the sub-classification method is preferable for at least three reasons (Inbens & Rubin, 2015). First, if the propensity score is mis-specified, it may generate large weights for observations whose propensity score is very close to 0 or 1. In the sub-classification method, propensity scores are implicitly smoothed within strata when the effects are averaged across all strata. Second, the estimates of the sub-classification method tend to have smaller variance because they are not affected by large weight observations as in the weighting method. Finally, the covariate adjustment performed by sub-classification is superior to the covariate adjustment by weighting because in sub-classification the coefficients of the regression function are allowed to vary within each stratum while the weighting estimator employs a single regression function for all the sample data.

Sensitivity analysis

Unconfoundedness cannot be tested. However, assessments can be made based on the data that make the hypothesis of uncounfoundedness less credible and plausible. If the data fail these assessments, then our assumptions about our ability of removing bias by adjusting for differences in the covariates' distributions of the project and control group need to be revised.

Estimating impact on pseudo-outcomes

The first approach consists of estimating the impact of the intervention on outcomes that are known not to be affected by the intervention. Consider dividing the set of covariates X in potential outcomes Xp and remaining Xr covariates. This assessment consists of estimating project effects on the Xp covariates using the Xr covariates for the estimation of the propensity score and for regression adjustment. One limitation of this approach is that it does not test all aspects of the conditional independence restriction but only the differences in the averages of the potential outcomes. There are two ways to make the assessment more restrictive. The first consists of subdividing the analysis of the outcome variables on intervals of the variable. For example, for a binary variable we could test that the impact on the potential outcome at different quintiles of the distribution are jointly 0. The second consists of testing the equality of the averages of the potential outcomes in sub-populations of the treatment and the control groups. Of course, in both cases, as the number of tests increases, the chances of finding a statistically significant difference also increases, and this should be taken into account in the analysis.

One difficulty in implementing this approach is that it is difficult to imagine a covariate that is not affected by the programme over time. All the covariates used in the propensity score and all the covariates in the potential pool of covariates consist of socioeconomic variables or demographic variables for which it is easy to imagine mechanisms through which they could be affected by the programme. Hence, using baseline-endline changes in these covariates as pseudo-outcomes is not feasible. The only exception is the household being affected by a weather shock such as a flood or drought. However, the project may change the impact of weather shocks and their perception. In addition, these events affect many households in localised areas and the impact on these events may not be efficiently adjusted by additional covariates.

The ideal setting for this type of assessment is when lagged values of the outcome variables are available in the covariate set. Our data contain lagged values of some variables based on retrospective questions. In particular, attendance rates can be calculated in relation to the 12 months before the survey as well as in relation to the previous 12 months. The baseline survey also collected retrospective data for several income components in relation to one year and two years before the survey. In particular, retrospective data were collected for a sub-sample of livestock holding, a sub-sample of agricultural crops, micro-enterprise profits and wages. The four income components were not collected at a level of detail that makes them comparable to the full income data. In our application we look at three pseudo-impacts: the impact on primary net attendance ratios one year before the survey; the impact on the value of livestock holdings over two years before the survey; and the impact on micro-enterprise profits one year before the survey.

Table A11 Pseudo-impacts of the MV project

Outcome	Coefficient	s.e.	P-value
Net attendance ratio (primary)	0.01	0.02	0.427
Value of livestock holdings	-17.09	136.88	0.901
Micro-enterprise profits	5.66	5.70	0.321

The estimation of the project impact on pseudo-outcomes reveals no surprises. None of the differences found is large in size and none is statistically significant.

Assessing project effects on pseudo-treatments

The second approach consists in testing the impact of interventions that are not implemented. The easiest way to conduct this assessment is by focusing on sub-groups of the control groups that do not receive the intervention. Consider three groups: project, control 1 and control 2, G includes (p, c1, c2). Units in G=c1,c2 receive the control treatment (P=0), while units in G=p receive the project treatment (P=1). We define group unconfoundedness as the group being independent of the outcomes given the covariates:

G _||_ Y1, Y0|X

Group unconfoundedness implies the testable restriction:

G _||_ Y|X, G includes (c1, c2)

The ideal set-up for this assessment is to have two control groups that are systematically different and likely to show biases in comparisons that are not adjusted by the covariates. Since none of the two control groups received the treatment, the project should have no impact on the outcomes of the two groups. Finding no evidence of any difference between the two groups does not imply unconfoundedness but makes it more plausible.

In our application, there is an obvious sub-division of the control group, which consists of the far away control villages in the Builsa and West Mamprusi districts. Villages located in the vicinity of the project areas cannot be used in this exercise because they are potentially affected by the intervention. On the other hand, far away villages are sufficiently distant from the project areas to rule out a significant impact of the intervention on most outcomes. At the same time, the two far away control regions belong to two different districts that are different in many ways and likely to exhibit biases in treatment-control comparisons.

In our application we look at three key outcomes: net attendance ratios in primary school; per capita expenditure; and per capita income.

Table A12 Impact of pseudo-interventions

Outcome	Coefficient	s.e.	P-value
Net attendance ratio (primary)	0.04	0.04	0.342
Per adult equivalent expenditure	-0.02	0.12	0.873
Per capita income	-0.38*	0.20	0.051

Table A13 Full income table of results

Outcome	Coefficient	s.e.	P-value
Average DD effect	-0.38*	0.20	0.051
DD effect first year	-0.42*	0.25	0.091
DD effect second year	-0.25	0.22	0.267
DD effect third year	-0.54**	0.27	0.042

Appendix B: Comparison of individual-level and village-level impact estimates

Most evaluations estimate project effects by drawing samples of equal size from each village and without adjusting the estimate by cluster population size using weights. The MV evaluation drew samples proportional to cluster size thus implicitly weighing the village-level estimates by the population size of each village. The impression is that sampling of equal size from each village is conducted more as a matter of habit and convenience than as a matter of choice, but the use of either approach responds to different evaluation questions. In some cases, we want to generalise estimated impacts of an intervention to a population of locations. For example, what is the impact of MV on other MV sites in Ghana? In other cases, we may want to generalise the estimated impact to a population of individuals. For example, what is the impact of MV on the population and on another MV population where the project is implemented? The two estimates can be identical and the difference is subtle but in some cases it can have large implications.

The impact of MV on a person (i) in site (j) is (Raudenbush & Bloom, 2015):

$$B_{ij} = Y_{ij}(1) - Y_{ij}(0)$$

The average impact of MV in village (j) is:

$$B_j = \sum_{i=1}^{N_j} \frac{B_{ij}}{N_j}$$

The average impact of the average village impact is:

$$B_{village} = \sum_{j=1}^{J} \frac{B_j}{J}$$

On the other hand, the impact of MV on people is the average village impact weighted by the population size of each village:

$$B_{people} = \frac{\sum_{j=1}^{j} N_j B_j}{\sum_{j=1}^{J} N_j}$$

If the impact of MV is the same in all villages, then the average impact across villages and the average impact across people are identical. However, if the impact of MV is correlated with village size, the two impacts may differ significantly. There are good reasons to believe that impacts of development interventions are inversely correlated with village size. Many interventions consist of forming groups or providing basic services or infrastructure. This is rarely performed proportionally to the target population. Smaller villages end up being disproportionately favoured by interventions and the potential benefits in larger villages are comparatively smaller. This means we would normally expect the average impact across villages to be larger than the average impact across individuals, because in the estimation of impacts more weight is given to villages of smaller size. Since most project effects assessed in impact evaluations estimate the average village impact, interpretations of these effects as impacts on individuals tend to overestimate the true population effect.

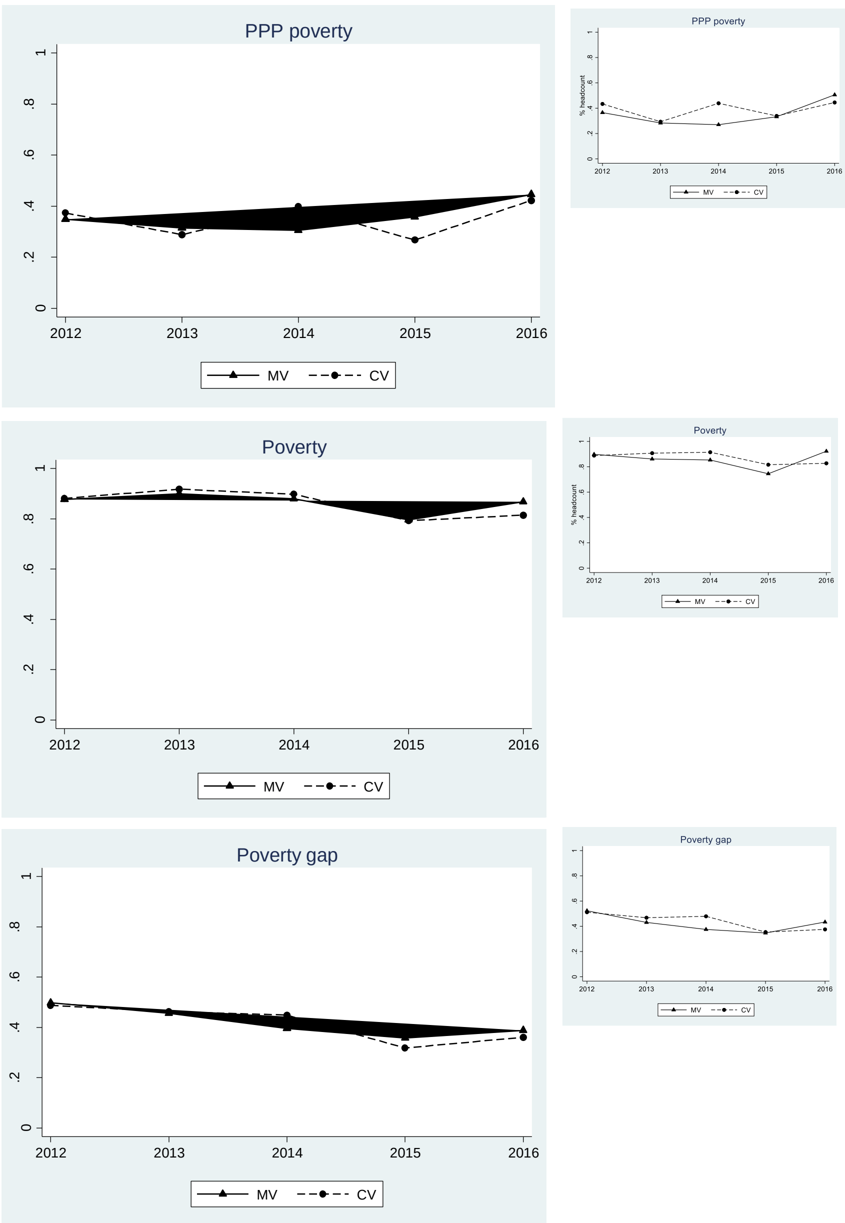
The population impact of MV can be obtained either by drawing samples that are proportional to sample size or by applying the population weights above to the analysis. In our application we drew samples proportional to village population size and we were therefore able to estimate the impact of MV in the population. If we now wish to estimate the village-level impact of MV we should undo the proportional weighting process. The MV impact at the village level is a weighted average of the impacts at the individual level:

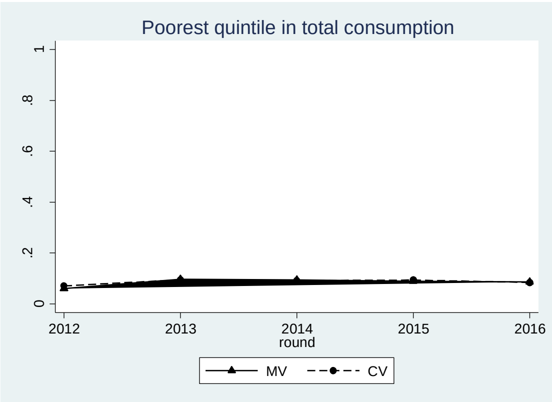
$$MV_{village} = \sum_{i=1}^{S} MV_i \frac{S}{Js_j}$$

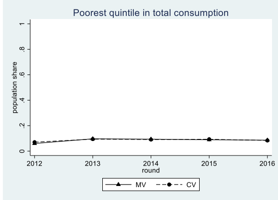
where S is total sample size and sj is the sample size in cluster j. In this way the cluster whose sample size sj is equal to the sample divided by the number of clusters S/J has a weight of 1. Smaller villages will have larger weights and larger villages will have smaller weights.

It might look odd to re-estimate the project effects at the village level when the population level effects can be calculated by design. However, we decided to do this exercise for two reasons. First, in this way we obtain an approximation of the impact we would have observed had we sampled a fixed number of individuals from each cluster without using weights at the analysis stage, as is normally done in similar impact evaluations. Second, any observed differences between the impacts estimated at the village and the individual level suggest there are impacts of the intervention that vary with the size of the villages, meaning the intervention is more focused or more effective in large or small villages.

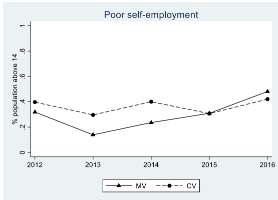
Figure B1 Individual- and village-level effects compared

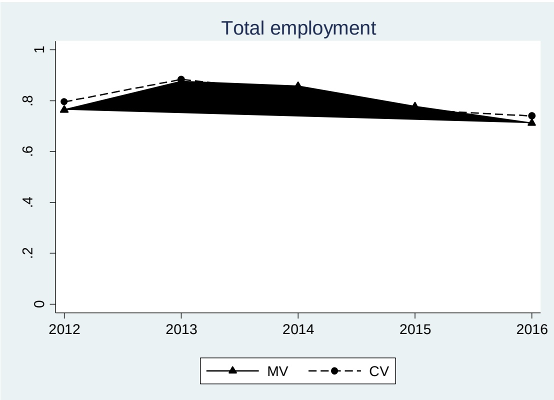


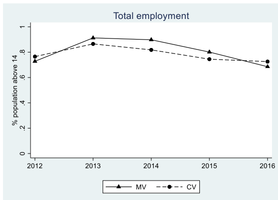

Poorest quintile in total consumption
0
.2
.4
.6
.8
1
2012
2013
2014
2015
2016
round
MV
CV

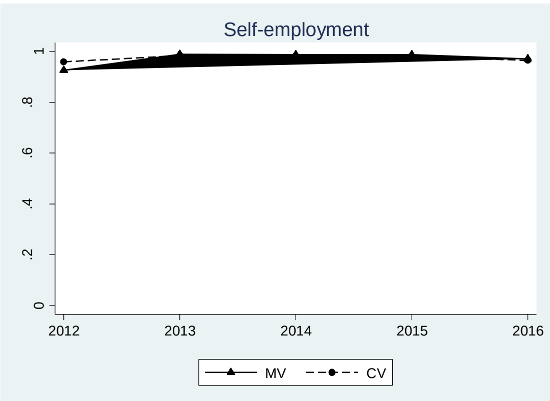

Poorest quintile in total consumption
population share
2012
2013
2014
2015
2016
round
MV
CV


Poor self-employment
0
.2
.4
.6
.8
1
2012
2013
2014
2015
2016
MV
CV

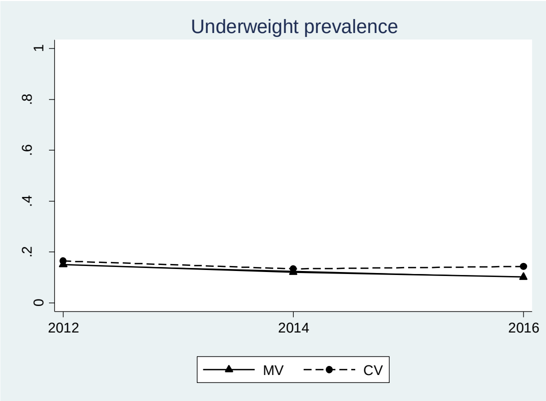

Poor self-employment
% population above 14
2012
2013
2014
2015
2016
MV
CV

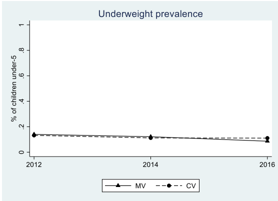

Total employment
0
.2
.4
.6
.8
1
2012
2013
2014
2015
2016
MV
CV

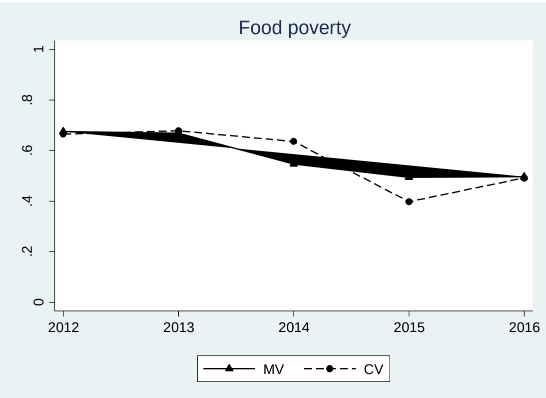

Total employment
% population above 14
2012
2013
2014
2015
2016
MV
CV

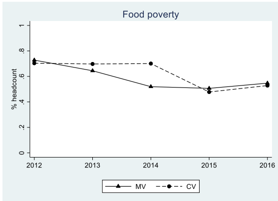

Self-employment
0
.2
.4
.6
.8
1
2012
2013
2014
2015
2016
MV
CV


Self-employment
% population above 14
2012
2013
2014
2015
2016
MV
CV









For brevity, we conducted the comparison between individual-level and village-level impacts of MV on the indicators of MDG 1, eradicate extreme poverty and hunger (Figure B1). Impacts of MV on poverty are visibly larger at village level than at individual level, though the differences are not statistically significant (see Table B1). Impacts on employment are also larger, while there is no difference in the impact on the fraction of underweight children. These data suggest that the impact of the intervention on economic outcomes is larger in small villages, perhaps because farmers' groups, savings groups and fertiliser are provided to each village and non-proportionally to population size. The village-level distributional effect of MV is also larger as both the poverty gap and the share of expenditure of the poorest quintile in the population decrease more using village-level estimates than using individual-level estimates. This is what we would expect because the interventions benefit a proportionally larger population in small villages than in large villages, and therefore make the population more homogenous in small villages in comparison with large villages.

Table B1 Individual- and village-level effects compared: MDG 1, eradicate extreme poverty and hunger

	Individual-level effects		Village-level effects	
MDG	**Baseline CV**	**DD average impact**	**Baseline CV**	**DD average impact**
Proportion of population below $1 (PPP) per day	37.28	4.32 (0.256)	43.37	4.85 (0.289)
Proportion of population below national poverty line	88.08	1.17 (0.676)	86.83	-0.64 (0.823)
Poverty gap ratio	48.72	-0.38 (0.869)	51.15	-1.60 (0.583)
Share of poorest quintile in national consumption	7.24	0.87 (0.321)	8.76	-0.76 (0.658)
Employment to population ratio	79.49	2.83 (0.252)	76.53	6.03 (0.085)
Proportion of employed people living below $1 (PPP) per day	33.52	0.36 (0.930)	39.96	4.64 (0.048)
Proportion of own account and contributing family workers in total employment	95.86	3.77 (0.059)	96.67	1.80 (0.685)

	Individual-level effects		Village-level effects	
Percentage of underweight children under 5	16.43	-0.51 (0.821)	13.69	0.29 (0.915)
Proportion of population below minimum level of dietary energy consumption	66.48	-0.55 (0.885)	70.35	-4.30 (0.435)

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economic Letters, 80*(1), 123-129.

Alkire, S., & Santos, M. E. (2010). Acute multidimensional poverty: A new index for developing countries. *OPHI Working Paper No. 38 QEH Oxford University*.

Alkire, S., & Santos, M. E. (2014). Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index. *World Development, 59*, 251-274.

Atkinsons, A. (2015). *Inequality: What can be done?* Cambridge: Harvar University Press.

Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Phylosophical Society, 78*(4), 551-572.

Benjamin-Chung, J., Abedin, J., Berger, D., Clark, A., Falcao, L., Jimenez, V., . . . Colford, J. (2015). *The identification and measurement of health-related spillovers in impact evaluations: a systematic review*. London: International Initiative for Impact Evaluation (3ie).

Bourguignon, F. (2011). Non-anonymous growth incidence curves, income mobility and social welfare dominance. *Journal of Ecnomic Inequality, 9*(4), 605-627.

Bourguignon, F., & Chakravarty, S. (2003). The Measurement of Multidimensional Poverty. *The Journal of Economic Inequality, 1*(1), 25-49.

Collet, D. (2002). *Modelling Binary Data*. London: Chapman & Hall/CRC.

Deaton, A. (1997). *The Analysis of Household Surveys*. Washington DC: The World Bank.

Deaton, A., & Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. *LSMS Working Paper 135, World Bank*.

Duclos, J.-Y., & Tiberti, L. (2016). Mulitidimensional Poverty Indices: A Critical Assessment. In M. D. Adler & M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press.

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

Fewster, R. M. (2009). A Simple Explanation of Benford's Law. *The American Statistician, 63*(1), 26-32.

Fields, G., Chichello, P., Freeije, S., Menendez, M., & Newhouse, D. (2003). For Richer or for Poorer: Evidence from Indonesia, South Africa, Spain and Venezuela. *Journal of Economic Inequality, 1*(1), 67-99.

Ghana Statistical Service. (2014a). *Ghana Living Standard Survey Round 6 (GLSS 6), Main Report*. Accra: Ghana Statistical Sevice.

Ghana Statistical Service. (2014b). *Poverty Profile in Ghana (2005-2013)*. Accra: Ghana Statistical Service.

Ghana Statistical Service, Ghana Health Service (GHS), & ICF International. (2015). *Ghana Demographic and Health Survey 2014*. Rockville, Maryland, USA: GSS, GHS, and ICF International.

Greene, W. (2011). *Econometric Analysis 7th edition*. Harlow: Pearson Education Limited.

Halliday, t. J. (2005). Mismeasured Housheold Size and its Implication for the Identification of Economies of Scale. *Oxford Bulletin of Economics and Statistics, 72*(2), 246-262.

Hill, P. (1986). *Development Economics on Trial: the Anthropological Case for a Prosecution*. New York: Cambridge University Press.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Impact Evaluation. *Journal of Economic Literature, 47*(1), 5-86.

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

Judge, G., & Schechter, L. (2009). Detecting Problems in Survey Data Using Benford's Law. *Journal of Human Resources, 4*(1), 1-24.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

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., Acharya, A., Barnett, C., & Dogbe, T. (2013). An impact evaluation design for the Millennium Villages Project in Northern Ghana. *Journal of Development Effectiveness, 5*(2).

Mosley, W. H., & Chen, L. C. (1984). An Analytical Framework for the Study of Child Survival in Developing Countries. *Population and Development Review, 10*(Suppl.), 25-45.

Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric Models and Economic Forecasts*. New York: Irwin/McGraw-Hill.

Raudenbush, S. W., & Bloom, H. S. (2015). Learning About and From a Distribution of Program Impacts Using Multisite Trials. *American Journal of Evaluation, 36*(4), 475-499.

Ravallion, M. (2012). Mashup Indices of Development. *World Bank Research Observer, 27*(1), 1-32.

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

Ravallion, M., & Chen, S. H. (2003). Measuring pro-poor growth. *Economics Letters, 78*(1), 93-99.

Rosenbaum, P. 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.

Sen, A. (1992). *Inequality Reexamined*. Cambridge: Harvard University Press.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*: www.stiglitz-sen-fitoussi.fr.

UNDP. (2010). *Human Development Report*. New York: Oxford University Press.

White, H., & Masset, E. (2003). The Importance of Household Size and Composition: An Illustration from Vietnam. *Development and Change, 34*(1), 105-126.