MILLENNIUM VILLAGES IMPACT EVALUATION, BASELINE SUMMARY REPORT

Date: **February 2014**

By Masset, Jupp, Korboe, Dogbe, and Barnett

itad

Results in development

IDS Institute of Development Studies

Participatory Development Associates Ltd.

London School of Hygiene & Tropical Medicine

Acknowledgements

This report has been prepared by the team for the impact evaluation of the Millennium Villages Project. The team is composed of staff from Itad, the Institute of Development Studies, the London School of Hygiene and Tropical Medicine, and PDA-Ghana. The team is fully independent of the Earth Institute and the Millennium Promise. The principal authors of this report are Dr Edoardo Masset, Dr Dee Jupp, Dr David Korboe, Tony Dogbe, and Dr Chris Barnett. The team is nonetheless very grateful to all the researchers that have assisted with data collection, the staff at DFID, and everyone else that has provided support, information, and comments – including the work of the Earth Institute during the enumeration phase. The findings of this report are the full responsibility of the authors, and any views contained in this report do not necessarily represent those of DFID or of the people consulted. The first drafts of this report were edited and proofread by Pippa Lord, Jane Stanton, Alice Parsons, and Kelsy Nelson. The final copy was proofread by Caitlin McCann.

Citation

Masset, E., Jupp, D., Korboe, D., Dogbe, T., & Barnett, C. 2014. *Millennium Villages Impact Evaluation, Baseline Summary Report*. Itad, Hove.

Acronyms

ANCOVA	Analysis of Covariance
CHPS	Community-based Health Planning and Services
CSPro	Census and Survey Processing System
CV	Control Village
DD	Difference-in-differences
DFID	Department for International Development
DHS	Demographic and Health Survey
EI	Earth Institute
FGD	Focus Group Discussion
GLSS5	Ghana Living Standards Survey 5+
GSS	Ghana Statistical Service
IDD	Initial Design Document
ISSER	Institute of Statistical, Social and Economic Research
JHS	Junior High School
MDGs	Millennium Development Goals
MICS	Multiple Indicator Cluster Survey
MMDA	Mamprugu-Moaduri District
MWDA	West Mamprusi District
MV	Millennium Village
MVP	Millennium Villages Project
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
PRA	Participatory Rural Appraisal
PRG	Peer Review Group
PVA	Poverty and Vulnerability Assessment
QA	Quality Assurance
RCA	Reality Check Approach
SADA	Savannah Accelerated Development Authority
SHS	Senior High School
TBA	Traditional Birth Attendant
UN	United Nations

Table of Contents

Tables

Figures

Boxes

1. Introduction

1. This report presents the baseline findings from the Department for International Development (DFID)-commissioned impact evaluation of the Millennium Village Project (MVP) in Northern Ghana.[1] The project will run from 2012 until 2016, with interventions targeting a cluster of communities with a total population of approximately 27,000 people. The MVP has been designed to demonstrate how an integrated approach to community-led development can translate the international Millennium Development Goals (MDGs) into results. It is an approach that has been previously piloted in Kenya and Ethiopia and in 2006 launched at scale to reach nearly half a million people across 10 countries throughout Sub-Saharan Africa. The new Millennium Village (MV) in Northern Ghana is the first to be accompanied by an independent impact evaluation. Details of the conceptual approach and methodology for the evaluation are presented in the Initial Design Document (IDD), with appendices containing the tools used for data collection.[2]

2. The evaluation uses a mixed methods approach to impact evaluation. At the core of the methodology is a difference-in-differences (DD) design that compares changes in outcomes in the MVP areas before implementation to post-implementation, with changes in the same outcomes for an explicit control group. DD allows the evaluation to isolate the MVP impact on outcomes (including poverty, child development, undernutrition, and child mortality) from effects of other variables changing over time.[3] Alongside the quantitative survey data, there are a number of supporting qualitative approaches that aim to better understand how and why change has occurred. There are four key qualitative methods. First, a **Poverty and Vulnerability Assessment** (PVA) describes local and multi-dimensional perspectives of wealth and well-being. Second, an **Institutional Assessment** captures empowerment and institutional change, particularly between the community and district levels. Third, a **Reality Check Approach** (RCA) uses a mini-anthropological study to better understand how the MVP affects the realities of people as well as captures any unintended consequences. And lastly, an **Interpretative Lens** approach takes the preliminary quantitative survey findings and obtains local feedback and interpretation around emerging themes of analysis. This last module will be deployed during the mid and end terms, when there will be a time series dataset that can provide a quantitative measure of change/impact. All other qualitative modules were deployed during the baseline.

[1] The Terms of Reference for the assignment is included in the Initial Design Document, Appendix A.
[2] Available at: www.ids.ac.uk/project/millennium-villages-in-northern-ghana-impact-evaluation
[3] While the implementation of a randomised trial is in principle possible by, for example, randomly allocating the interventions to matched village pairs, it would have been highly impractical in this particular case and the cost would have been prohibitive. The matching of control villages to project villages (on aggregate characteristics) and further matching of project and control households at the analysis stage (on household characteristics) within a DD approach appears to be the next best feasible approach after a randomised design. For more detail, see Masset et al. (2013).

3. The purpose of this document is to:

- Present the research instruments, both quantitative and qualitative, employed by the evaluation team
- Provide a description of the MV and Control Village (CV) communities
- Describe the main characteristics of the study population in terms of MDGs, poverty, and other outcome indicators
- Test the validity of the control group
- Assess the quality of the data collected

4. Section 2 provides an overview of the MVP area and some background information on its ethnic mix, geography, physical infrastructure, transport, and the local economy. Section 3 gives an overview of the survey instruments and qualitative research. Section 4 presents the characteristics of the study population, covering the baseline findings for MDG status, household characteristics, poverty, agriculture, health, education, gender, social networks, and attitudes towards risk. Section 5 then provides a summary of the characteristics of the data, focusing on balancing tests, the issues around seasonality, and an assessment of the data quality. Section 6 provides the summary, with conclusions about the characteristics of the study population as well as the baseline dataset (the balance tests, seasonality issues, and data quality).

5. This Baseline Report has been submitted to a two-stage review process. Firstly, the Evaluation Advisory Group (EAG), which is made up of stakeholders from DFID, EI and government, met in August 2013 to provide comments on the first draft. The purpose of the EAG is to ensure the continued relevance of the evaluation to a broader set of policymakers, as well as to contribute particular insights about how the MVP and other interventions in the region are operating on the ground. The comments of the EAG are advisory only, and while the evaluation team is obliged to justify its response, it is not required to alter the report to obtain approval. At the second stage, the Peer Review Group (PRG) reviewed the technical quality of the report in October/November 2013. The comments of the PRG have been satisfactorily addressed prior to the publication of this report. The PRG is being coordinated by the International Initiative for Impact Evaluation (3ie) and includes the following international experts:

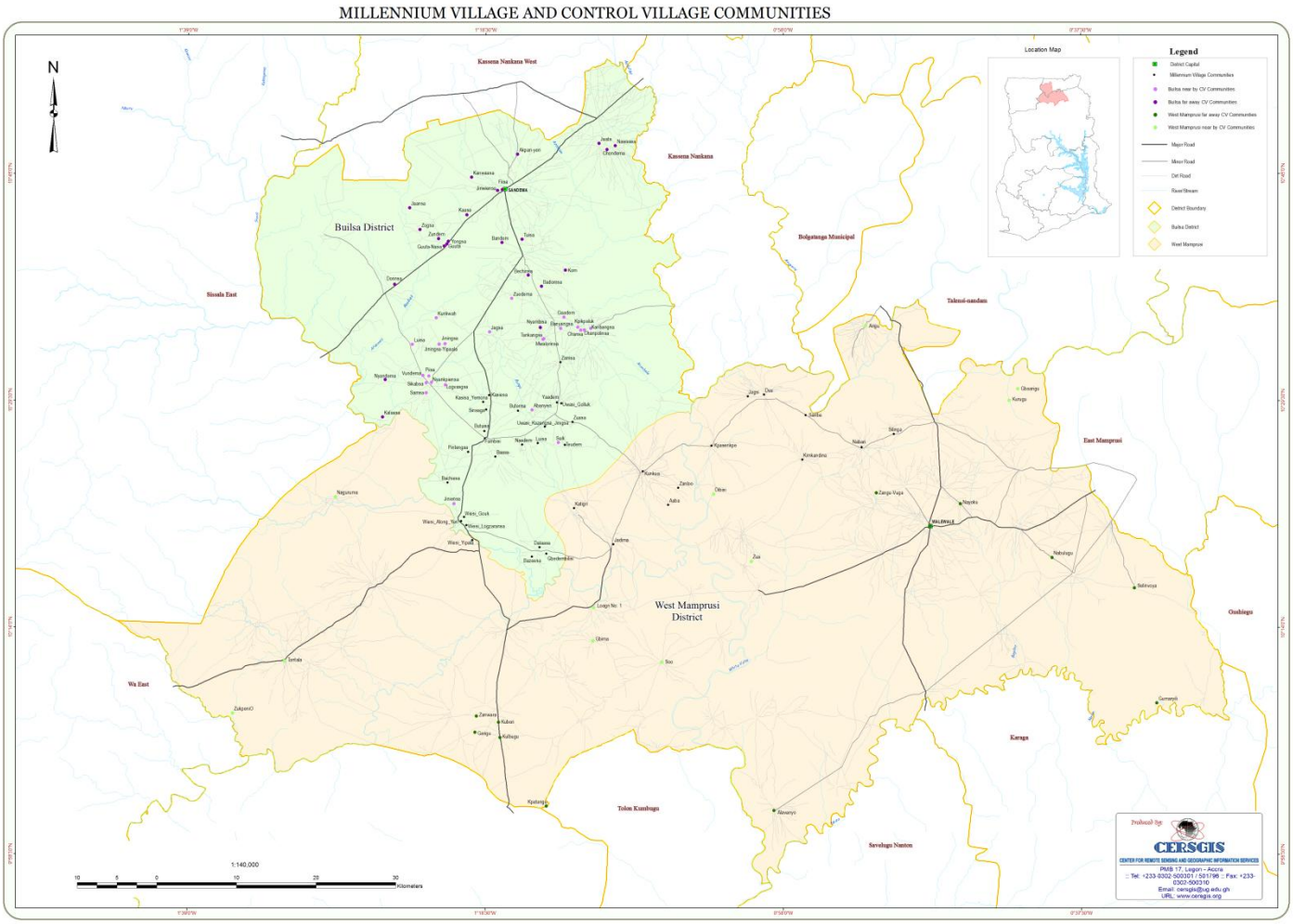
- Howard White, Executive Director, 3ie
- Annette Brown, Deputy Director, 3ie
- Thomas de Hoop, Research and Evaluation Specialist at American Institutes for Research
- Chris Udry, Professor of Economics, Yale University
- Robert Osei, University of Ghana

6. **Note:** As this is an ongoing study at the time of the public release of this report, the names of control villages and the identities of particular focus groups have been replaced with identification references to protect the identity of the control communities and the study population. For individuals and households, these are and will not in any case be presented in the data analysis and the evaluation findings.

2. Overview of the MVP area

7. This section provides an overview of the MVP area and is based on observations during reconnaissance visits to the Savannah Accelerated Development Authority (SADA) MVP sites, information gathered from the Reality Check, and focus group discussions (FGDs) with community members and district officials. The MVP area (including the control sites) is located in the former districts of West Mamprusi and Builsa in Northern Ghana. The community groups forming the MV cluster span across two different districts and have different languages and customs.[4] In late 2012, these districts were split to form the four districts of **West Mamprusi** (District Assembly based in Walewale), **Mamprugu-Moaduri** (District Assembly based in Yagaba), **Builsa North** (District Assembly based in Sandema), and **Builsa South** (District Assembly based in Fumbisi).

Figure 1. The MVP and Control Communities



[4] A detailed discussion of the selection process for the MVP sites is provided in the IDD, pages 25-27.

2.1 Institutional assessment

8. When the MVP began operations, it was split over the West Mamprusi and Builsa Districts. In 2012, the two districts were divided into four: West Mamprusi (old district), Mamprugu-Moaduri (new district), Builsa South (new district), and Builsa North (old district). The MV communities are in the first three districts with control communities located in all four districts. As of the end of 2012, the newly created districts did not have functioning offices though some district staff had been appointed and were working at their posts. For this reason, the leadership of the two old districts thought, for the institutional assessment, it would be more prudent to bring the staff of the new and old districts together for an FGD. The former were new to the district and were yet to acquire or settle into their offices. Furthermore, the MVP had been dealing with the staff of the old districts.

9. To understand the institutional arrangements the districts have put in place for the implementation, management, and tracking of the effectiveness and sustainability of the MVP, the administrative and department heads engaged in an FGD as each group plays a different role and experiences a project differently. These discussions took place on 16 November 2012 in the Builsa North and South Districts and on 22 February 2013 in the West Mamprusi and Mamprugu-Moaduri Districts at their respective district assembly halls. From the discussion, the focal persons in the districts' administrative arm for the MVP interventions are the District Planning Officers. The agriculture, education, health, works, and cooperatives each have focal people who participate in meetings organised by the MVP. The representatives of the Departments of Community Development and the Department of Social Welfare in the two districts appeared to have little knowledge of the MVP interventions because they did not directly participate in the meetings organised by the MVP. The representatives of the Departments of Community Development felt that considering their links with communities they should have been more involved than they were. The District Planning Officers who participated in the discussion did not have a copy of the final MVP project document nor did the representatives of the other departments. The MVP secretariat organises quarterly stakeholder meetings and workshops with the core administrative and decentralised department staff to provide updates on activities and progress on interventions.

10. According to the administrative leadership of the two old districts, they were involved in the formulation and design of the project and approved of intervention areas, namely agriculture, health, education, and infrastructure. However, they said they were not involved in the selection of the project communities, which they would have preferred since they are more conversant with their districts. In West Mamprusi, the district stakeholders believed that as a consequence of the selected project communities being located along the main road, the off-track and 'deeper overseas' communities where poverty and deprivation is very severe are ignored.

11. In both districts, department representatives working directly with the MVP felt involved in the planning and implementation of interventions such as the provision of boreholes and training of health volunteers, but felt they had limited roles in the monitoring and evaluation of the projects. Resource allocation to communities is done by the MVP and SADA and not by the districts, which the department representatives felt does not foster a sense of ownership by district officials despite their involvement in the planning and implementation.

12. The district assemblies contribute to the MVP by providing logistical assistance such as lending district vehicles, granting the use of the assembly hall for meetings with project stakeholders or assembly personnel, and technical assistance from district staff for project implementation.

13. Administrative and department heads explained that from previous experience, projects that were sustained long after they ended were effectively monitored to ensure the interventions reached the target communities. User fees on services or facilities like an ambulance service can be introduced at the end term when income levels have improved so that funds can be recouped for project maintenance. Also, project sustainability depends on the extent to which community members and departments are actively involved in the implementation so as to ensure that a sense of ownership of the project and enough capacity building has taken place. Sustainability can be secured by building solid and durable physical structures like schools and clinics that are disaster resistant.

14. Overall, the district officials anticipate that the MVP interventions will reduce poverty as income levels increase, that infrastructure facilities will be improved and relieve the district assemblies of the cost, and that the district staff capacity will be strengthened.

2.2 Communities in the area where MVP operates

15. The descriptions presented in this section of the report refer to the communities in the focus group component of the study. Unlike the communities in the two Mampruli-speaking districts, which are characterised by a dispersed settlement pattern, those in the Builli-speaking districts follow a nucleated settlement pattern and are close to each other. Communities in Builsa North and South are closer to their district capitals of Sandema and Fumbisi, respectively. With the exception of BSCF1 and BSMV5, the rest of the participatory rural appraisal (PRA) for communities surveyed by our study[5] in the Builsa South District are located within a distance of 5-20 kilometres of the district capital. Communities in the West Mamprusi and Mamprugu-Moaduri Districts such as MMCF2, MMCF1, MMCN2, and MMMV1 are more remote (over 20 kilometres to the district capital). The most remote community in the West Mamprusi District is MWCN1, a settler community with no social amenities. The area is subject to seasonal flooding, which leaves some communities cut off from others, and thus they are commonly referred to as 'overseas.'

16. **Ethnic mix.** Almost all of the communities are multi-ethnic as a result of migration, inter-tribal marriages, and diverse arrangements with settler populations such as the Fulanis. The ethnic groups in these communities differ according to area. The most common ethnic migrant groups are Kassena, Guruni, and Kantosi, plus the presence of Dagaaba and Busansi. Communities in the Builsa districts are predominantly Builli-speaking. Likewise, those in the Mamprusi districts are predominantly Mampruli-speaking. A significant exception is in MMMV1, a Builli-speaking community within the West Mamprusi District. With the exception of MMCF1, which is populated by Mamprulis exclusively, the rest of the communities have more than one ethnic group.

17. **Community organisation.** A number of the project and control communities in the Builsa North and South districts are regarded as sections of much larger communities. For this reason, these areas are traditionally under the leadership of sub-chiefs who report to the main chief of the larger community. Of the 16 communities in the Builsa North and South Districts selected for the qualitative baseline study, only three were not a section of a larger community. These are BNCN1, BNCF1, and BSMV4. The rest are regarded as sections.

[5] See Section 3 for a description of how we obtained a sample of study communities from the full sample of localities in the two districts of intervention.

18. The communities are structured along traditional lines with shared powers between chiefs and sub-chiefs on the one hand and between chiefs and the landowners on the other. Whilst the chiefs or sub-chiefs role generally relates to governance, the landowners are the 'spiritual caretakers' of the land and are responsible for allocating it.

19. The focus groups felt that the Assembly representatives, as individuals, listen to the community and can be trusted in local governance arrangements[6] whereas the District Assemblies are seen as detached and unresponsive. Even though Assembly representatives are not part of the traditional governance hierarchy, they play an important role in community development. They are consulted by chiefs and sub-chiefs on matters relating to the well-being of the community. Together with the unit committee, they are ranked high in terms of reliability, accessibility, and effectiveness in assisting community members with health, education, and sanitation issues.

20. The communities also have a number of existing self-help groups such as farmer groups, other occupation-based groups, women's groups, youth groups, and widows groups.

21. **Physical infrastructure.** None of the 21 communities selected for the focus group component of the impact evaluation have a tarred road linked to a major town such as district capital. However, the roads linking these communities are fairly motorable with the exception of those between MMMV1 to Djardema, BNCF1 to Sandema, and MMCF1 to MMCF2, which get flooded during the rainy season.

22. The most visible forms of physical infrastructure in these communities are public basic schools and Community-based Health Planning and Services (CHPS) compounds. All seven of the MV communities selected for the baseline study have schools. In some schools, the highest level of education taught is primary class six whilst in others it is Junior High School (JHS) three. Communities without a JHS are generally able to access one in a nearby community within approximately five kilometres. The communities that do not have schools are MMCF1, MWCN1, BNCF3, BSCN2, BSCF1, and BNCF2.

23. Another form of physical infrastructure is mechanised boreholes, mostly provided by the District Assemblies. As they are not equitably distributed by population size, communities with very few boreholes supplement their water supply with hand-dug wells and streams.

24. Except for BNCN1 and BNCF1, none of the communities surveyed by the study were connected to the national electricity grid. MWMV1, MMCN2, BSMV5, MMMV1, MMCF2, and BSCN2 have solar street lamps erected at vantage locations such as CHPS compounds.

25. **Transportation.** Although the roads linking some of the communities to the district capital are fairly motorable, transport to and from these communities is a major challenge due to the region's poor vehicular system. The primary means of transport is cycling and motorbikes. The newly introduced tricycles called 'motorkings' are owned by a minority of households and are used for commercial purposes.

26. In some communities, transport to markets is only available on market days via trucks. Communities such as MMCF1 and MWCN1 have no public transport to market centres and thus people must walk more than 10 kilometres to access public transport.

[6] Note: the RCA did not concur with this finding.

27. Communities such as MMCF1, BSMV5, MMCN2, and BNCF1 lie within the flood zone and are cut off from major markets and district capitals during the rainy season. With the recent introduction of 'motorkings' or 'motor kia,' movement to district capitals and other major towns on non-market days has improved.

28. **Economic profile.** The main economic activities in these communities are farming and trading in agricultural produce. Almost every household has a farm, which is primarily used for subsistence consumption. A small proportion of the harvest is sold to cover other household costs such as education, health, and migration.

29. Market days are very important in the economic setup of the Builsa and Mampruli districts. They are also a major cause of girls' and boys' irregular school attendance in communities close to the markets. Apart from the major markets in Sandema, Fumbisi, and Walewale, there are additional satellite markets in Kaadema, MMMV1, and other non-MVP communities in Wiaga and Wulugu.

30. The communities lack basic economic endowments such as access to information, agro-processing, and manufacturing industries. All the communities studied are agrarian and survive on rainfed agriculture.

31. Besides farming, other economic activities include picking wild fruits and shea nuts, oil extraction, and producing charcoal. There is an active labour force ready to work, but these people are forced to take up employment outside of the communities due to a lack of opportunities. Non-farming seasons are plagued with idleness across both project and control communities. Over the past three decades, migration has become an important coping strategy. Migration generally occurs during the dry season and income generated is used to finance farm inputs, education, healthcare, and trading.

32. These communities have been recipients of a limited number of development initiatives from both government and non-governmental organisations including World Vision, Presbyterian Agricultural Station, Technoserve, and SADA, with interventions mostly targeted at reducing vulnerability, poverty, and improving resilience to disasters.

Comparison between MV and Control Communities

33. According to the MV project managers, one factor in selecting project communities was that they had to be in clusters.[7] This seems to account for the proximity of the project communities to each other and major routes. Four of the MVs out of the seven studied are located along the major road to the district capitals or markets whilst their corresponding control near and far communities are remotely situated. This means that they have less direct access to major marketing centres in the district.

34. Based on the village narratives compiled for the 21 communities (seven MVP, seven control near, and seven control far) subject to the qualitative components, it seems the MVs in the qualitative study are slightly better off in terms of infrastructure than the CVs. For education, there are schools in six out of seven MVs of which two do not have a full set of classrooms from pre-school to JHS 3. In the control, there are schools in five out of the seven near communities. Of the five, only three are from pre-school to JHS 3. In the other two, one goes up to pre-school and children have to access education three kilometres away. In another community, the school is under construction after a community initiative. In the far CVs there are schools in three out of the seven communities. In CVs

[7] For details of selection process please see pp. 25-29 in the IDD.

without schools, children have to travel on average about three kilometres to access one. In one of the three CVs the school classes only go up to class six. There is a similar distribution of health facilities to education in the areas. For the communities in the qualitative baseline study, there are health facilities in four out of seven MVs and two out of seven in control far and two of seven in control near communities. In terms of access to electricity, there is not a great difference between the MVs studied and the CVs. Only one MV and one control near and one control far are connected to the national electricity grid. Three MVs, one near CV, and one far CV have solar lamps located in parts of the community. In regards to water access, once again there is not much difference between the MVs and CVs. Six out of the seven MVs studied have boreholes with hand pumps. There are also hand-dug wells constructed by the community members. Five of the control near and five of the control far communities studied have boreholes with hand pumps.

35. After the new districts were created in 2012, subsections of some communities were legislated as separate communities. There are more instances of this in the Builsa South District. As a result, three out of five of the MVs studied are full standing communities. In the same district, the two remaining MVs studied and four near CVs as well as five far CVs are all sub-sections of larger communities. This occurred less in the West Mamprusi District (MWDA) and Mamprugu-Moaduri District (MMDA). The MVs and the far and near CVs are all whole communities, with the exception of one near CV, which is a subset of a larger community in these districts. These changes have implications for community organisation. The sub-set communities that are now full communities are traditionally under the supervision of a sub-chief who reports to the main chief of the larger community. Whilst this may not be a problem immediately, in the future a new generation of these sub-set communities that become full communities may struggle with their parent community for autonomy and access to resources.

36. The major livelihood in both project and control communities studied are farming and livestock rearing. Farming is done by households in all well-being categories. Women in both MVs and CVs also participate in trading. Women in the MVs that are on major transport routes are into much larger trading than those in their control near and control far communities that are located remotely with very difficult access to transportation.

3. Survey instruments and qualitative research

37. This section summarises the quantitative data collection undertaken for the baseline, including the processes of supervision and quality control. The section also summarises the qualitative evaluation modules, and ends with a discussion of the challenges of applying a theory-based impact evaluation approach to the Millennium Villages Project.

Survey data collection

38. As detailed in the IDD, the evaluation adopts a DD design by comparing the change in outcomes in the MVP areas before implementation to post-implementation, with changes in the same outcomes for an explicit control group. MV communities were matched with CV communities using community-level characteristics summarised by a propensity score. Two stratifications were adopted: district and geographic distance from the MV areas. Each project community was paired with a *nearby* control community and a *faraway* control community, and the pairing was conducted separately in the Builsa and West Mamprusi districts.[8]

39. Prior to the quantitative data collection, a census listing (the 'household count') was conducted in 35 project communities and 68 matched control communities.[9] In the MV areas, 750 households were randomly selected from the household listing proportionally to the village population size. Similarly, 750 households were selected proportionally to the village population size in the nearby CV areas and another 750 in the faraway CV areas. The survey teams were only able to interview 712 of the selected 750 households in the project areas and 1,466 of the originally selected 1,500 households in the control areas. The total sample size of the survey therefore stands at 2,178 households, of which 32.7% resides in the project areas. This means that about 18% of the households listed in the census were eventually interviewed. The Earth Institute (EI) did not employ a replacement protocol, which resulted in the loss of a fraction of the planned sample. This fraction is very small because the MVP survey team visited each target household up to three times when it was not found. The overall response rate was 96.8% (94.9% in the MV areas, 98.4% in the near CV, and 97.1% in the faraway CV). Reasons for households not being found or interviewed were not reported during the data processing. However this information will be added by the MVP to follow-up datasets as it was collected during follow-up interviews in cases where it was missing. Table 1 provides a summary of the start and end dates for each survey instrument.

[8] Note: At the time of selecting and pairing the control communities, there were two districts covered by the MVP (Builsa and West Mamprusi). These districts were later split to form four districts although the area covered is the same.

[9] The final number of control communities is larger than twice the number of project communities because in the matching process a project community of about 10 households was considered to be part of another community.

Table 1. Actual enumeration start and end dates

	MV		Builsa Control Sites		West Mamprusi Control Sites	
	Start	End	Start	End	Start	End
Household Count	30 January	4 February	5 June	16 June	26 June	4 July
Demographic Collections	19 March	26 March	28 June	7 July	20 July	29 July
Adult Survey	16 April	21 June	1 August	13 August	25 August	18 Sept.
Household Survey	7 May	28 June	17 August	18 Sept.	1 Sept.	18 Sept.
Anthropometric Collections	10 May	17 June	29 August	20 Sept.	26 August	20 Sept.
Blood Survey	9 May	10 June	29 August	20 Sept.	15 Sept.	26 Sept.

Source: EI Report, SADA Ghana Survey Collection Dates and Response Rates.

40. Households were administered a number of survey instruments in order to track progress on the MDGs and other indicators selected by the evaluation team. Some of these instruments were designed by the MVP based on previous MVP sites and others by the evaluation team in collaboration with the MVP. All instruments were reviewed and approved by the evaluation's Peer Review Group (PRG) to ensure rigour and maintain independence. The full list of survey instruments with the numbers of interviews conducted is presented in Table 2.

Table 2. Quantitative survey instruments and number of observations

	Interviews MV areas	Interviews CV areas	Total interviews
Community and facilities questionnaire	35	68	103
Household questionnaire	712	1,466	2,178
Adult female questionnaire (15 to 49)	848	1,995	2,843
Adult male questionnaire (15 to 49)	504	1,129	1,663
Anthropometry (under 5)	608	1,353	1,961
Blood tests: anaemia & malaria (under 5)	381	409	790
Cognitive tests (children 6 to 19)	1,160	2,296	3,456
Easy education tests (children 9 to 19)	573	1,163	1,736
Advanced education tests (children 11 to 19)	111	299	410
Time preferences and income expectations	432	704	1,136

Quality control of survey data

41. This section provides an overview of the data collection process and the quality checks undertaken for the 2012 baseline. The section summarises the checks undertaken by the MVP, as well as additional checks undertaken by the independent evaluation team. There are three main parts to the MVP's quality control system: (i) *Field-based systems that are used for capturing enumeration errors* including detailed form checks that take place three times for each questionnaire (once by the enumerator, a second time by a data editor, and third by a field supervisor); (ii) *Random spot-checks* of enumerators as conducted by field supervisors (with protocols for the random survey spot-checks); and, (iii) Once the field checks are complete, questionnaires are sent to the field office for *single entry, double data entry, and cleaning in CSPro* (Census and Survey Processing System).[10] The

[10] See: http://www.census.gov/ipc/www/cspro/

MVP deployed a total of 36 enumerators and eight editors and supervisors. For the control sites household survey there were a total of 39 enumerators and eight editors and supervisors.

42. **Field-based and random checks to capture enumeration errors.** The MVP's field-based systems make use of quality control forms that are used to verify the accuracy of the survey data. Different forms are used for the Household Survey and the Adult (Male/Female) Surveys, and in each case the supervisor selects a household (or individual, in the case of the Adult Survey) at random for verification. The supervisor will then re-interview the main respondent/individual, asking them a series of short, non-sensitive, and easily verifiable questions (10-15 minutes). The responses are recorded and then compared to the completed questionnaire previously filled out by enumerators. This is undertaken during the first 10 days of enumeration, where at least one household per day (or two female respondents per day, in the case of the Adult Survey) is selected randomly for re-enumeration/verification visits, from the target list of households enumerated that day. When selecting households/respondents, supervisors are to ensure that every enumerator had at least one revisit during the verification period. The same process is repeated during the last 10 days of enumeration.

Table 3. Re-enumeration/Verification Checks conducted by the MVP

	Target for verification	Actual verified for MV Surveys	Actual verified for Builsa Surveys	Actual verified for WM Surveys
Household survey quality control form	At least one household per day for re-enumeration/verification visits during the first 10 days and the last 10 days of enumeration. A minimum of 20 household questionnaires (about 5-10% of the original sample).	53	69	50
Adult female quality control form	At least two respondents for per day for re-enumeration/verification visits during the first 10 days and the last 10 days of enumeration. A minimum of 40 female adult questionnaires (constituting about 5-10% of the original sample).	24	58	62

43. **Data entry and cleaning in CSPro.** Questionnaires are single entered by a data clerk, cleaned, and then sent to a different data clerk for double entry, followed by another round of cleaning. The data entry templates and cleaning scripts contain four checks: (i) Missing data, (ii) Invalid response codes, (iii) Logical/consistency, and (iv) Structural. An overview of the data entry and cleaning systems using the education modules as an example is provided in Table 4.

Table 4. Types of data entry checks

Data entry checks	Description	Example (from the Education module)
Missing	Check if the required question is answered.	Q133 (ever attended school) cannot be left blank.
Range	Check if the response is within the allowable range of responses.	Q133 has the response codes of: "1 - yes"; "2 - no"; and "90 - Don't know." Any response not equal to one of these values is considered out of range.
Logical/ consistency	Check if a particular response is consistent with a previous response.	If the individual has never attended school (Q133=2), then the rest of the questions in the module should be skipped.
Structural	Check for duplicates, that structural relationships hold within survey modules, and all household members are accounted for across all modules.	The name and ID of the individuals listed in the education table are consistent with the name and ID of individuals in the demographic/member roster.

44. The survey data (Household and Adult) are subject to a double entry system so that data on a number of key variables are re-entered for verification purposes in CSPro. The first cleaning stage involves the application of pre-established structural, range, logic, and consistency checks. The advanced cleaning stage involves an additional layer of quality/analytical checks[11] (e.g. outlier checks, continuous/open-ended data). Standardisation focuses on value labels, the missing and 'N/A' Codes, and Advanced Structural Checks. The de-identification internal release process strips names and other direct identifiers from the data in addition to scrambling IDs.

45. Independent Quality Checks by the Evaluation Team. **The independent evaluation team undertook additional checks to independently verify the effectiveness of the MVP's data collection and quality assurance system. The work of the external quality assurance (QA) team focused on the following:**

- **Enumeration and Supervisor Shadowing** processes to check the quality of training and fieldwork, including adherence to the MVP's quality control processes. These QA checks include the shadowing of enumerators and supervisors during enumeration.
- **The Process Check** tool, which was used by the external QA team to check that the MVP's processes (according to the EI Enumeration Manual[12]) were correctly followed.
- **Random spot-checks of Households and Adults**. This involved revisiting a sub-sample of households (random selection of 5% of the sample) to check: (i) The household members can be located and existed; and, (ii) Responses to a selection of questions within the survey instruments to test errors as a result of expectation bias or measurement inaccuracy.

46. A summary of the checks is listed in Table 5.

[11] At the time of writing, the EI's own report of their QA processes was not available. We would expect this to include details of the actual cleaning process, and whether data are 'corrected' or 'discarded' and on what basis.
[12] The Earth Institute, Columbia University, Survey Enumeration Manual – Guidelines for enumerators, field supervisors, and data managers, January 2012.

Table 5. Summary of independent quality checks

	MV Site	Control Sites (Near and Far)
	Numbers of Checks	**Numbers of Checks**
Training Observation	Field team attended training organised by EI on 10-12/4/2012, 23-30/4/2012	Training attended for nine days, starting on 17/9/2012
Supervisor Process Checks	4 – with 2 supervisor checks in Builsa and 2 in West Mamprusi Districts	4 conducted – with 2 in Builsa and 2 in West Mamprusi Districts
Comprehensive Process Checks	1 conducted	1 conducted
Shadow Enumeration	22 adults and 27 HH were shadowed during the enumeration process	None conducted – since no significant errors were identified from the shadowing exercise in the MV site, it was decided to discontinue these checks in the CV sites and focus resources on the spot checks
Spot Checks – Adults	25 spot checks – 14 female and 11 male	97 adult females – at least one female from each household that was checked – were spot checked
Spot Checks – HH	31 spot checks were carried out in relation to the HH survey	37 HH in near control site and 37 in far control site = 74 HH in all

Qualitative evaluation modules

47. The quantitative work was complemented by three strands of qualitative research. The first comprised of a series of FGDs with residents grouped by gender and well-being in 21 communities (seven MVs, seven nearby CVs, and seven faraway CVs). In total, 84 such focus groups and an additional 24 school-based interviews were conducted. In a separate strand employing the RCA, a team of researchers lived with selected households (both MV and CV) for short periods in order to ground truth and qualify the findings of the quantitative survey. And lastly, an institutional assessment was undertaken with the four districts' authorities. A total of 112 FGDs were conducted.

48. The aim of the qualitative work is to complement the quantitative work by shedding light on how and why things have changed, particularly from the perspective of community members and with a special focus on the experience of disadvantaged and marginalised groups. The qualitative methodologies are detailed in the IDD[13] and in Appendices C – G of this report. Below is a brief summary:

- **Poverty and Vulnerability Assessment.** During the baseline, this work focused on identifying categories of households by different wealth/well-being groups and enabling community members to define poverty in their own terms by developing key indicators of wealth/well-being at the community and household levels. The well-being categorisation exercise was undertaken in seven project communities as well as 14 corresponding far and near control communities in the two (now four) districts. This process was the basis for forming focus groups determined by both the well-being

[13] See IDD Appendix E, Part 1 (Quantitative Data Collection Instruments) and Part 2 (Qualitative Methods and Tools).

categories and gender, which were then used to gather baseline information on poverty, vulnerability, and local institutional issues. These focus groups will be used in subsequent years to consider changes since the baseline and capture local interpretations of the emerging quantitative findings (i.e. the Interpretative Lens approach). The work on well-being categorisation took place from 12th to 30th November 2012, with the focus group fieldwork during January to February 2013.

- **Reality Check Approach.** The RCA is a series of mini-anthropological studies used to better understand how the MVP affects 'ordinary' people as framed by their own realities at the individual and household levels. It also aims to capture unexpected effects of project implementation. RCA is a proven methodology undertaken in four countries to date, and draws on similar work conducted in another six countries. It is based on the principles of immersion, using participant observation and the conversations that take place during a four-night stay in the homes of poor families. The training of the team and pilot took place in December 2012, with the fieldwork conducted between February and March 2013.

- **Institutional Assessment**. The institutional assessment is based on a series of FGDs and interviews with district staff, undertaken in November 2012 and February 2013. During the baseline, the two original districts were both split into two, while some staff of the two new districts had been newly recruited, and many did not have functioning offices. In the light of this, the study team organised FGDs by bringing the two split districts together. The focus groups were separated between the core staff of the district administration[14] and the heads of departments and agencies. This is because they play different roles in the MVP and in projects in general and hence experience the projects differently. The institutional assessment is complemented with ongoing data collection of district expenditure and time use in the MV and CV areas.

Applying a theory-based approach

49. At the core of the Millennium Villages impact evaluation is a difference-in-differences design based on a statistical analysis of the quantitative dataset. Alongside this is a mix of other methods, drawing on theory-based impact evaluation (TBIE) approach to evaluation (White 2009). TBIEs seek to elaborate the programme theory in order to better explain the impact (net effect). Therefore, while the quasi or experimental designs focus primarily on measuring the impact (what has changed), the TBIE approach is used to open up the 'black box' to answer questions about why an intervention has achieved its intended impact and how it worked (or otherwise). The aim is to yield evidence about how the programme is working, rather than just if it is working.

50. It is important to note that there is no single 'theory of change' for the MVP as a whole, and indeed there is no definitive example being used by the project implementation team. This is partly because of the sheer complexity of the MVP (in terms of the number and sequencing of interventions), but also because the MVP seeks to implement a 'learning-by-doing' approach. Based on interactions with the Earth Institute, the evaluation team have identified three possible sources for a more orthodox 'theory of change':

    - Firstly, an overarching 'theory of change' based on the economic theory of the 'poverty trap' and grounded in theoretical and empirical research. This will in any case be tested through data being collected by the evaluation (for a full discussion see the IDD, pages 10-16).

[14] Namely, the District Chief Executive, the District Planning Officer, the District Coordinating Director, and the District Budget Officer.

- Secondly, there exists a series of detailed, generic 'intervention logics' that show the anticipated interconnections from inputs-to-outputs-to-outcomes, and then to MDG-level impacts (see IDD, Appendix B). It is important to note however that these are not specific to the northern Ghana MVP, and make reference to interventions that will/may not occur. Plus, despite their apparent detail (they have been summarised into thirteen A3 pages for the IDD), they contain many unknowns and many imprecise connections (particularly in terms of the sequencing and synergies between different sectoral activities).

- Thirdly, there is the 'logframe,' which attempts to fit the MVP into a linear hierarchy from activities through to impact. While this is the only example of a type of 'theory of change' that is based on the specifics of the northern Ghana MVP, it has significant drawbacks in terms of being a highly simplified/stylised version of the underlying logic. In particular, it does not adequately reflect the complexity of the MVP operation, and does not provide sufficient detail on the assumptions and connections between outputs, outcomes, and impacts.

51. While the 'intervention logic' (second bullet, above) seems closest to a more usual 'theory of change' (i.e. mapping the causal logic, assumptions, etc.), it does not provide a robust basis for framing the quantitative/qualitative analysis. This is for a number of reasons, including: (a) It is a generic intervention logic that is non-specific to the MVP being implemented in northern Ghana; (b) It is too lengthy and cumbersome to provide a meaningful framework for identifying key areas within the 'black box'; and, (c) It is still too vague about specific activities/interventions, how they are to be sequenced, and, importantly, where the synergies lie. In short, any attempt to utilise this version of the MVP's generic 'intervention logic' as the basis for a theory of change risks leading to an overly mechanistic approach to the qualitative work (i.e. long lists of questions/areas of enquiry; too little focus on picking up unexpected findings; insufficient use of open-ended areas of enquiry; etc.).

52. In conclusion, the challenge of applying a TBIE approach to a complicated programme such as MVP has led the evaluation team to further refine its approach to one that does not assume a single (all encompassing) theory of change. The evaluation team proposes to:

- Firstly, at the mid/end term stages, the data collection is to be sequenced so that some of the emerging quantitative findings can be further explored through the qualitative work (the Interpretation Lens approach, as outlined in the IDD). This will ensure that there is a clear connection between the two datasets, and specifically, it will allow the qualitative modules to directly contribute to the interpretation of emerging quantitative findings. This process will also help ensure that the qualitative work is prioritised, with it being focused on those aspects of the theory of change where explanatory/qualitative information is most useful.

- Secondly, the evaluation team will further develop an overarching 'theory of change' that focuses on the core institutional issues and their contribution to longer-term sustained change. This will draw upon both the literature and the emerging qualitative findings from the Reality Checks, the focus group discussions, and the institutional assessment. This institutional framing will be used to help guide the work of the Institutional Assessment, in particular.

- And lastly, the evaluation team will develop an initial set of 'micro theories of change' that focus on selected parts of the MVP intervention logic – and which need further exploration through the mid/end term data collection and analysis (e.g. seed and fertiliser supplies). This will draw together the literature review, field observations, monitoring data from EI/SADA, as well as discussions with project managers. The number of 'micro theories of change' will necessarily be

limited – in part due to resource constraints but also because focusing on the minutiae of the causal chain for specific activities/sectors tends to downplay the integrated/synergistic dynamics of the MVP.

53. The above still aims to satisfy the two principle objectives of a taking a TBIE approach: (i) A focus on understanding how and why the impacts of the MVP have occurred; and, (ii) Providing sufficient integration between the qualitative and quantitative data collection and analysis to allow the core impact questions to be answered in a robust manner.

4. Characteristics of the study population[15]

54. This section sets out the characteristics of the study population by focusing on the MDG status in the MVs and CVs, local perceptions of poverty, and household characteristics. The section then considers different aspects of people's lives including income poverty, agriculture, education, and health in addition to gender relations, social networks, and people's preferences and expectations.

4.1 MDG status in MV sites and comparison to the rest of Ghana

55. The goal of the MVP is improving the MDGs. The MDGs are presented in the MV and CV areas separately as they emerge from the baseline quantitative data in Table 6. For comparative purposes, the same indicators are shown for three other areas: the *rural north* (comprising the rural areas of the Northern Region, the Upper East, and the Upper West), *rural Ghana* (comprising all rural areas in the country), and *Ghana* (national level data). Since the rural north is the poorest geographical region in the country and the project selected an extremely poor cluster of villages for the intervention, the indicators should improve as Table 6 is read from left to right.

56. The indicators were calculated following the instructions from the official United Nations (UN) handbook for monitoring progress on MDGs. In the case of poverty and employment indicators (the first panel in Table 6) figures were calculated from the Ghana Living Standard Survey 5 (GLSS5) data of the Ghana Statistical Service (GSS) whilst all other figures were calculated using the Multiple Indicator Cluster Survey (MICS) data. Whilst looking at the indicators in Table 6 it should be kept in mind that the MVP survey is not strictly comparable to the other datasets because they were collected at different times, with different sample sizes, and employing different questions.[16] Note also that because of the size of the MVP survey and of the characteristics of the questionnaires, not all MDG indicators can be calculated and are therefore not reported in Table 6.[17]

[15] All tests of statistical difference between samples reported in the tables of this section and throughout the report are employing cluster standard errors following the methodology presented in Section 5.

[16] GLSS5 is a nationally representative living standard survey conducted between September 2005 and September 2006. Expenditure data were collected year-round and with the supervised use of expenditure diaries rather than by a standard survey questionnaire. There is disagreement about whether expenditures from diaries are more or less accurate than expenditures from survey questions but they do deliver different figures. The sample size consists of 8,700 households. The MICS data are a nationally representative survey modelled to the DHS surveys. Data were collected on 12,150 households between September and December of 2011. Because a large portion of the EI survey was also modelled to the DHS survey, there is a high degree of similarity between our datasets and the MICS data.

[17] The indicators excluded for lack of data are: Growth rate of GDP per person employed; Proportion of population below minimum level of dietary energy; Proportion of pupils starting grade 1 who reach last grade of primary; Share of women in wage employment in the non-agricultural sector; Proportion of seats held by women in national parliament; Maternal mortality ratio; Unmet need for family planning; HIV prevalence among population aged 15-24 years; Condom use at last high-risk sex; Ratio of school attendance of orphans to school attendance of non-orphans; Proportion of population with advanced HIV infection with access to treatment; Incidence and death rates associated with malaria; Incidence, prevalence and death rates associated with tuberculosis; Proportion of tuberculosis cases detected and cured directly; Internet users per 100 population.

Table 6. MDGs in MV and CV localities and in the rest of Ghana

Indicator	Millennium Villages (MV)	Control Villages (CV)	Rural Northern Ghana	Rural Ghana	Ghana
Goal 1 Eradicate extreme poverty and hunger					
Proportion of population below $1.25 (PPP) per day	74.0	71.9	68.5	39.3	28.5
Poverty gap ratio	35.6	33.7	31.6	13.5	9.6
Share of poorest quintile in national consumption	4.6	5.6	2.5	4.9	5.2
Employment-to-population ratio	76.6	79.5	63.6	58.7	68.8
Proportion of employed people living below $1 (PPP) per day	52.7	51.8	65.1	36.8	25.8
Proportion of own-account and contributing family workers in total	98.0	98.9	96.1	88.0	75.4
Prevalence of underweight	16.5	14.3	23.6	16.5	14.2
Prevalence of stunting	26.8	29.6	37.5	28.7	24.6
Goal 2 Achieve universal primary education					
Net attendance ratio in primary school	60.5	68.9*	59.0	66.4	71.2
Net attendance ratio in middle school	9.7	15.4	11.7	22.2	30.6
Literacy rate of 15-24 year olds	33.1	33.8	30.0	37.7	37.9
Goal 3 Promote gender equality and empower women					
Ratios of girls to boys in primary school	1.26	1.03	1.03	1.01	1.03
Ratios of girls to boys in middle school (JHS)	1.40	1.74	1.45	1.31	1.27
Goal 4 Reduce child mortality					
Under-5 mortality rate	70.6	105.1**	103.30	87.7	81.1 [b]
Infant mortality rate	43.2	69.9**	59.4	55.9	52.7 [b]
Proportion of children immunised against measles	86.2	75.1**	89.6	88.0	89.7
Goal 5 Improve maternal health					
Proportion of births attended by skilled health personnel	28.7	29.4	41.0	55.6	69.4
Contraceptive prevalence rate	10.0	10.3	16.6	27.6	29.5
Antenatal care coverage	96.6	85.4***	95.4	96.5	97.1
Goal 6 Combat HIV/AIDS, malaria, and other diseases					
Proportion of population with comprehensive correct knowledge of HIV/AIDS	16.0	17.3	14.6	16.7	22.6
Proportion of children under 5 sleeping under insecticide-treated bed nets	32.5	54.0***	48.3	48.3	41.1
Proportion of children with fever treated with anti-malarial drugs	29.9	38.3	42.5	41.4	42.6
Goal 7 Ensure environmental sustainability					
Proportion of population using an improved drinking water source	73.2	72.2	72.5	74.2	86.0
Proportion of population using an improved sanitation facility	8.7	8.7	8.9	47.9	66.0
Goal 8 Global partnership for development					
Telephone lines per 100 population	0.0	0.0	0.2	0.3	2.3
Cellular subscribers per 100 population	58.7	49.0**	50.1	68.1	80.0

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

57. Are the MVs worse off than the CVs? The MV localities were targeted based on an assessment of the levels of deprivation and there is a possibility that they are structurally worse off than other localities in the vicinity. Table 6 shows that there are differences between the MV and CV areas and of which some are statistically significant. Some of these differences can be the result of small sample sizes and of chance error. Other differences, like the availability of mosquito bed nets, are likely to be the result of seasonal factors (see Section 5.4), whilst others can be the result of structural differences between the two areas. The emerging pattern appears to be the following: First, there are no differences between the MV and CV sites in terms of poverty levels, characteristics of employment, and other indicators of material living standard such as access to water and sanitation facilities (though households in the MVs appear to have more mobile phones). Second, households in the MVs are better off in regards to health indicators. The MVs display lower child mortality rates, higher immunisation coverage, and antenatal care. Third, children in the MVs appear to be less educated. Primary school attendance is slightly higher in the CVs though there are no differences in literacy rates and gender parity in schooling. These differences in health and education indicators suggest that there might be some underlying structural differences between the two areas in terms of access to education and health services, whether provided by the government or NGOs.

58. Are households in the MVs worse off than households living in the rural north? The data in Table 6 suggest this is not the case. The MV localities appear to be quite representative of general living conditions in the rural north as there are no large differences. Poverty rates and employment characteristics are very similar as well as education indicators of school attendance, literacy, and gender parity. There are some differences in health indicators and thus the picture is more mixed: children from the MVs appear to be better off in terms of undernourishment and mortality rates, but the provision of health services, such as malaria treatments and assistance at delivery, is higher in the rural north.

59. Are MV localities worse off than the rest of rural Ghana? This is where large differences emerge. Poverty rates in the rural areas are from 2006 and therefore not strictly comparable to those calculated based on the MVP survey but the MV areas appear to be much poorer than rural Ghana. Access to sanitation facilities is also much more common in rural Ghana as well as the provision of health services, such as malaria treatment and assistance at delivery. No major differences emerge with respect to enrolment, literacy rates, undernourishment, or mortality rates.

60. Are MV localities worse off than the rest of Ghana? The differences are really visible in answering this question. Income poverty and inequality are much higher in the study area than in the rest of Ghana. This is not a surprise as the intervention area was selected because of its high level of deprivation. Perhaps more surprising is that there are no differences in terms of the nutritional status of children as measured by the prevalence of underweight children under five. An analysis of the causes and severity of malnutrition needs to go beyond the indicators adopted by the UN on the MDG list to include the distribution of outcomes in the population and seasonal stress in addition to the general composition of diet and micronutrients intake.[18] All employment indicators show that in the study area most people are engaged in a large number of economic activities compared to the rest of the country, which is probably explained by the large number of low quality occupations or because all

[18] A more detailed analysis of nutrition data and of iron deficiency is conducted in Section 4.7 and further work in this area will be conducted at the analysis stage.

household members have to do some form of work to survive.[19] Education attainment is lower in the study area and the quality of education is doubtful as demonstrated by very low literacy rates obtained from adults' reading tests. In terms of access to education, there appear to be no gender differences as there are larger proportions of girls attending school than boys at each educational level. In some of the FGDs, girls reported that they chose to stay in school rather than work as *kaya yei* in the streets of the more prosperous south.[20] Following two decades of sustained development efforts encouraging households to educate their daughters, fewer parents are now willing to send them to live with urban-based relatives as 'foster children,' a euphemism for unpaid housework. Overall, the FGDs were still ambiguous about girls' educational progress. Whilst the RCA confirms that in several schools girls outnumber boys, especially at the primary level, it seems this is because boys are more likely to drop out of school to work on the farm or herd animals rather than because more girls are enrolling than boys.

61. Maternal health indicators show a very low number of births attended by skilled professionals,[21] few antenatal visits, and low levels contraceptive use. Child mortality rates are comparable to those observed in all of Ghana. Knowledge of malaria is limited and the cases treated with orthodox drugs are fewer than in the rest of the country. Access to water and sanitation are similar to those observed at the national level. As can be expected, the use of landlines and mobile phones is well below the national average. The RCA indicated that although half of the houses had their own mobile phones, most other households could access them from their neighbours if they did not own phones themselves.

62. In conclusion, MV and CV areas appear fairly similar. There are differences, some of which can be the result of structural factors, such as the access to public or NGO provided education and health services. CVs display better education indicators whilst MVs show better health indicators. However neither area can be conclusively classified as better or worse-off than the other. There are no large differences between the MVs and the rest of the rural north. This suggests that the MVs might be considered representative of the wider rural north. However, there are large differences between the MVs, rural Ghana, and all Ghana. The data paint a picture of a deprived area where economic, education, and health conditions are very poor. The differences are particularly large in terms of monetary economic indicators of poverty. There are also some surprising facts. In particular, undernourishment and child mortality rates in the MVs are similar to those observed in the rest of the country.

4.2 Local perceptions of poverty[22]

63. Poverty is perceived by people not only in monetary terms but also in more nuanced ways. Among the main criteria that participants in the FGDs distinguished the poor from the rich were:

[19] The employment to population ratio is the proportion of individuals aged 15 or older who worked at any time during the 12 months preceding the interview either in formal or informal jobs. The ratio typically falls between 50% and 75% but is often higher than 80% in very poor countries reflecting a large number of low quality occupations or simply the fact that all household member have to do some work to survive.

[20] *Kaya yei* (singular: *kaya yoo*) are female head porters who sell their labour in urban market places. Recently, the term has been corrupted to include girls who wash pots and pans in street-side eateries.

[21] It needs to be noted that several of the traditional birth attendants (TBAs) met in the course of the RCA had received extensive training from NGO programmes and were not strictly 'unskilled' although the survey would count them as such. Currently, few TBAs attend deliveries, with most referring their clients to the formal health facilities instead as a result of a change in national health policy.

[22] This section is based on the wealth and well-being categorisation rankings, using participatory techniques to capture local perceptions of poverty.

- **Annual harvest volumes**: the well-off are those perceived to have year-round food security; this was the most dominant distinguishing metric
- The range and quantum of **physical assets**: the well-to-do have a wider range of capital assets (e.g. farm inputs and livestock) and in more substantial amounts
- Various **norms and cultural practices**: e.g. widows are largely considered to be poor, mainly because they tend to lack control over the assets of their deceased husbands in addition to having less decision-making power
- **Health and educational** status
- **Participation in community life**

64. This is illustrated in Tables 7 and 8.

Table 7. Local categorisation of well-being cohorts

Language group	Rich/well-to-do	Moderately rich	Poor	Very poor
	Bundan-tiri (rich person who has everything)			*Nambo* (extremely poor)
Builli (spatial area: Builsa)	*Jigsura* (wealthy person) *Dobroa* (weighty, heavy); at some sites (BNCF1 and BSMV2), variations include: *Ghantanyona* and *Pagroa/Pagra* (all meaning rich); at BSCN4 and BSCF1, the rich are referred to as: *Pagrim* (strong) or *Nyontanyona* (property owner)	*Azunchonga* or *Zunchong* (well-to-do person)	*Jajak* (poor person) In BNCF1, there is also *Jajakpiak* (harsh poverty); another variation in BSCN4 is *Nuwoba* (weak person)	*Nubowa*, *Akanuroa*; in BSCF2 and BSCN4, there is also *Jajak silinyieng* (no hope poverty, most used in reference to poor people who are disabled); Another expression used in BNCF1 is *Nuwobataaling*

65. The distinguishing features of the different well-being categories as defined by community members are indicated in Table 8.

Table 8. Characteristics of each well-being category

Rich/Well-to-do	Moderately Rich	Poor	Very poor
They are able to feed their household three times daily throughout the year. They live in houses built with blocks and roofed with zinc. They have large (five acres of land and beyond) farms and never get short of food throughout the year. They are respected in the community by both the people and the traditional leaders. They start farming immediately after the first rain, which signifies the beginning of the farming season, because they have money to acquire a tractor and also buy farm fertiliser. Their children are well fed and are constantly in good clothes. They sleep on mattresses. They have a 'bank account' i.e. livestock such as cattle, sheep, goats, and fowls are sold to raise money immediately during difficult times (more than six of cattle, sheep, goats, and fowls). They own motorbikes and bicycles and are able to transport their wives to the market on	They are able to provide three square meals to their family with little or nothing to spare. Some live in houses built with blocks and roofed with zinc. Some of their children ride bicycles to school. They have a lot of household labour and are able to hire a bullock or tractor to plough their farm. They own motorbikes and bicycles. They have livestock which they get money from immediately if they need money. They own electrical appliances such as TVs and radios. They have enough food to feed their household throughout the year because they are able to acquire fertiliser and tractors to plough the farms. They have cows (four and above), sheep, goat, and fowls. Their children also look healthy. They own motorbikes and the men give rides to their wives to	They live in mud houses. They possess very few or no livestock. They are unable to immediately rebuild their houses when destroyed by storms and mostly stay with relatives for a long time until they are able to rebuild. Their children do not go to school especially during the hungry season. In most cases, their young children (less than 10 years) also work to assist around the house. These children are normally used as farm labourers. They offer labour to those who are well-to-do to get money to purchase seeds and acquire the services of a bullock to plough their land. Their children are not well fed and always look sick and skinny. Their children are always in tattered clothes and walk barefoot even to farm. They depend on local herbs when they are sick because they cannot afford to go to the hospital. They do not possess either bicycles or motorbikes and always walk long	They are regarded as the worst in the community in terms of survival. They live in mud houses roofed with thatch. They have smaller farm sizes, i.e. about one acre, because they do not have enough labour to cultivate large farms. They use manual labour for farming throughout the season because they do not have the means to acquire bullocks, ploughs, or tractors. They depend on herbs when they are sick because they cannot afford to go to the hospital. Some of them, e.g. the blind, lepers, cripples, etc., cannot farm so they depend on the benevolence of other community and family members to survive. They always look sick due to the nutritionally poor food they eat. They command very little or no respect in the community. They sometimes drink a lot of alcohol.

Rich/Well-to-do	Moderately Rich	Poor	Very poor
market days. Their children ride bicycles to school. They have access to farm labour because they have money to hire the poor during farming season. Their farming activities are not affected even if their children migrate because they can hire people to farm for them. They own mobile phones, solar lights, and rechargeable lamps which their children use to do homework.	the market and sometimes their children to school. They are able to take their children to the hospital when they are sick because they can afford it. They have family members who have migrated to the south and send money to them on timely basis.	distances to farm. Their wives walk long distances to the market every market day. They are always unhappy and drink a lot of alcohol to help them forget their problems. They offer labour to the rich for survival. Their children sometimes look after the livestock of the rich and do not go to school. They farm on a small piece of land (one to two acres) because they cannot afford fertiliser and bullocks to plough their land. Some of the men within this category are not married. A widow with a lot of small children is also considered poor because their breadwinner is dead and they cannot farm to feed their children. Their children drop out of school very early. Their children normally migrate to places like Accra, Kumasi, Obuasi, Techiman, etc. to work for money mostly on vacation and sometimes when school is in session.	

66. The RCA findings and focus group discussions indicate that farming is, by far, the primary livelihood across MV and CV communities. Poor women and men are involved in subsistence farming of food crops as well as a small number of fowls and the occasional goat or sheep. Poor women's farms tend to be the smallest and are dominated by vegetables, legumes, and groundnuts that require relatively less labour and/or inputs. In contrast, men cultivate carbohydrates (maize, millet, and guinea corn), with legumes and pulses (typically groundnuts and bambara beans) as supplementary crops. In many cases, women do not have full control over the plots they farm as they are considered to be visitors regardless if they have married into the household. Whilst women are expected to help on men's farms with sowing, weeding, and harvesting, there is no reciprocal obligation for men to assist on women's farms.

67. Richer men tend to have the largest farms and keep larger ruminants. They hardly use their own labour but rather employ labour from other households. They have greater access to inputs such as land, seed, fertiliser, labour, tractors, and ploughs. Whilst the rich generally farm the same crops as the poor, their farms are bigger, diversified, and more dispersed. Rich men may also participate in trading as wholesalers and middlemen, dealing in cereals (maize, millet, and guinea corn), livestock, and occasionally agrochemicals (especially fertilisers and weed killer). In some of the Builsa communities, rich men also finance artisanal mining activities, employing the labour of poor young men as diggers in the mining pits. Rich women too may cultivate wholesale grains, shea nuts, and pulses or process and bulk trade in shea butter. Reflecting the high level of insecurity, the sector workers are perceived to be rich mainly because they have a steady income stream that enhances their access to food throughout the year.

68. Some of the poor men hunt game or harvest roofing thatch to sell to the rich men. Both the poor and rich produce charcoal though the rich have larger operations. They may also engage in by-day labour,[23] weave ropes from kenaf fibre, and produce baskets and *zaana*[24] mats from guinea corn stalks. Some poor women engage in retailing cereals, typically in meal-size portions. They also gather tama (shea nuts) to sell to the richer women while others sell their labour to the rich for farming, quarrying stones, plastering mud walls, and providing household chores. In some households, poor women rear fowls and the occasional small ruminant (mainly goats, but also sheep), only to be liquidated in an emergency. Though women's involvement in raising small ruminants is increasing, a married woman is still expected to seek her husband's consent before doing so. In the words of a poor woman in one FGD, "they [the men] own us and everything of ours." In spite of nutritional challenges faced by many households, domestic fowls are not routinely used for consumption by the poor. Whilst the rich may slaughter them for meat or on festive occasions, the poor generally save their livestock almost exclusively as security or for ritual sacrifices.

69. The rich tend to use agrochemicals whereas the poor may make use of animal droppings.[25] Because the poor have fewer animals, they have less access to opportunities that counter soil depletion. Even when subsidised fertiliser and tractor services are available, they tend to arrive too late in the farming cycle to be of use. Veterinary services are seen as important for the survival of livestock but are difficult to access, especially for the poor.

[23] Wage labour with payment typically made on the same day.
[24] Straw.
25 The RCA findings suggested that the poor have few ruminant animals, if any, and are not using any fertiliser (except green manure in some instances).

70. Poor households sometimes employ rotating pooled labour arrangements to maintain their farms and harvest crops. In the agricultural slack season, the practice is extended to include house construction.

71. Due to the increasingly volatile climate and over-exploitation of the same lands without any form of crop rotation, a range of crop yields are declining. In the Mampruli sites, groundnut yields have been badly affected, compelling many farmers to shift attention to cultivating beans. Builli communities also report that millet yields are dropping. However, their staple foods are so dependent on millet that they are unable to make a shift to other crops.

72. The increasing demand for cash to meet needs such as health insurance premiums, school supplies, farming inputs, dry-cell batteries, and lighting fuel compels farming households to sell a larger proportion of their harvests than in the past. The high cost of funerals, marriages, cyclic festivities, and sacrifices is another factor contributing to selling yields.

Difference in well-being categorisation between the MVP and control areas

73. Since the well-being categorisation was done with a mixed group of community members in each community, using their own criteria to determine which household belongs to which well-being category (and therefore perception based), it is not appropriate to compare one community with the other. Moreover, it is possible that in a few communities the members sometimes may have sought to give the impression that there were more poor households in the hope that more assistance will come to the community. The main value of the exercise is that in the subsequent rounds of studies where almost the same mixed group will use the same criteria from the baseline to determine which households have moved up or down the well-being ladder and the reasons for these changes. The focus in this work is therefore on trends and explanation rather than absolute measures of poverty.

74. Despite the caution above, if one makes a comparison between the MVs and the CVs, then in the Builsa South District the community members in the five MVs classified 17.5% as rich, 30% as averagely rich, 35% as poor, and 17% as very poor. In their corresponding control-near communities, 13% were classified as rich, 31% as averagely rich, 35% as poor, and 21% as very poor. In the control far communities, 11% were classified as rich, 46% as averagely rich, 24% as poor, and 19% as very poor households.

75. In the MMDA and MWDA, the mixed groups of community members in the two MVs on average classified 2% as rich, 25% as averagely rich, 51% as poor, and 21% as very poor. In the control near communities however, 22% were classified as rich, 64% as averagely rich, 9% as poor, and 5% as very poor. In the control far communities, 17% were classified as rich, 28% as averagely rich, 31.5% as poor, and 13% as extremely poor.

4.3 Household characteristics

76. The survey interviewed nuclear households and not extended households (Box 1). The average size of a household in the study population is seven members. Women head about 10% of households and about 20% are polygamous. There are more out-migrants than in-migrants in the area. About 1 household in 10 is hosting a migrant, whilst approximately 50% of households have a member who has temporarily migrated. The FGDs suggest that whilst nearly everyone who leaves the community intends to be away only temporarily, a significant minority end up being away for lengthy periods. Women are as likely to migrate as men. The main reasons for temporary migration is work (50% of cases) followed by live with or care for a family or friend (25%), and then education (18%).

Table 9. Household structure

	MV	CV
Household size	7.2	7.0
Female headed household	0.09	0.12
Polygamous	0.22	0.20
Average number of in-migrants per household	0.10	0.07
Average number of out-migrants per house	0.46	0.30*

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

77. The RCA and FGDs confirm the high rate of out-migration (mostly as seasonal farm workers and market porters), which is increasing due to needing cash for farm inputs, health, and education.

Box 1. Defining the household (restricted and unrestricted definitions)

Like other living standard surveys, the MVP survey considers a household to be a common decision making unit where all members share income and other resources and 'normally eat from the same pot.' A usual member of a household is a person who (whether present or absent at the time of the data collection) has spent at least the last six months in the household. But people in Northern Ghana often live in households sharing the same compound with other related households. Households or individuals in one compound can also be related to people living in other compounds in many different ways. This 'household' is considered to be the decision maker in most of the quantitative analysis and is also used as a denominator, for example, in the calculation of per capita expenditure. But to what extent is the 'household' considered by the quantitative survey to be a decision-making unit? Do households share a common budget? Do parents decide about the education of their children?

The qualitative studies suggest a more complex situation, and quite a wide variation in what constitutes a household. The RCA found it easier to refer to the entire compound built around a single courtyard as one household. It revealed that some compounds comprise different generations and may include widowed or abandoned sisters who have returned to live in the ancestral family home, grandchildren living with grandparents, nieces and nephews with aunts and uncles, or several siblings living with their dependents in one compound. The men, even of different generations, tend to make economic decisions together but men and women of the nuclear families living within the compound make the decisions related to their own families, such as education and health, themselves (though the FGDs found that women are often sidelined even in decision-making at this nuclear level). Clearly, the conventional definition of 'normally eat[ing] from the same pot' applies less to the reality of compounds where cooking is sometimes shared, family members eat at different times of the day, or where no cooking is done at all on some days. The qualitative studies raise issues around the extent to which respondents understood the survey's definition of 'household' and to what extent they answered consistently as a household or as a compound.

The quantitative study and the Reality Checks use different definitions of household, both of which have advantages and disadvantages. For this reason, the interpretation of the quantitative findings needs to be carefully considered and the Reality Check provides a counterpoint to understanding the full complexity on the ground.

4.4 Income Poverty

78. Over the last 20 years poverty reduction efforts have been substantial in Ghana, particularly in urban areas (World Bank 2011). Northern Ghana however has not been as positively affected by the economic growth and as a result poverty levels today are much higher than in the rest of the country. Table 10 provides the data on poverty, calculated using per capita expenditure data from the MVP survey which includes both cash expenditure and consumption of self produced goods and gifts.[26] The calculations are based on a poverty line of $1.25 per person per day after applying an adjustment for purchasing power parity (PPP) to per capita expenditure figures. Poverty rates are very high in the MV villages (74%) but not significantly higher than in the CV areas, and a bit surprisingly, inequality indicators are higher than in the rest of the country. Despite the casual impression of undifferentiated poverty, there is large inequality of incomes, at least on a per capita expenditure basis, within the MV area.

Table 10. Poverty indicators

	Poverty headcount	Poverty gap	Squared poverty gap	Gini coefficient
MV villages	74.0	35.6	21.6	0.44
CV villages	71.9	33.7	19.7	0.41
SADA region[a] (2006)	58.3	24.9	13.6	0.43
All Ghana[a] (2006)	28.5	9.6	4.6	0.42

[a] Data from the GSS living standard survey of 2005/2006.

79. The MVs are largely food-based subsistence economies. The majority of household expenditure (both cash and consumption of home produced goods) is on food (74%) of which at least half is home produced. The RCA revealed that families at the time of the study[27] spent almost nothing on food, surviving entirely on their own stored harvests of maize flour and beans, supplemented by leaves gathered nearby. The only food expenditures were on seasonings. The data display unusual food Engel curves that increase with per capita expenditure suggesting that any additional income is spent on food because of the high levels of deprivation. However, a closer analysis of measurement error in expenditure data suggest that this is not the case and that the food Engel curve has the usual shape of decreasing food shares as income increases.[28]

80. After food, the largest expenditure is on personal care, which includes items such as soap, hair dressing, personal care, and cleaning products. Overall, education and health are negligible components of total expenditure. However, in terms of proportion of cash expenditure for the household these become significant costs (particularly secondary education which requires lump sum payments). The RCA suggests that the increased need for cash, in what had been until recently a largely cashless society, has fuelled the search for paid work outside the community and is beginning to reduce the willingness of households to participate in traditional reciprocal labour arrangements.

[26] Details on the calculation of per-capita expenditure figures are reported in Appendix C.
[27] February to April 2013.
[28] See the analysis in Appendix I.

Table 11. Per capita expenditure and food shares

	MV	CV
Per capita expenditure ($PPP)	401	406
Food share	0.74	0.72
Share of own produced food	0.51	0.50
Clothing	0.01	0.02
Personal care	0.08	0.08
Health	0.01	0.02**
Education	0.01	0.01
Transport	0.03	0.03
Fuel	0.03	0.04**
Durables	0.04	0.03**

*Difference statistically significant at 10%, ** significant at 5%, ***significant at 1%

81. Despite a large proportion of imputed expenditure devoted to food, households are not able to meet their food requirements (Table 12). More than 80% of households reported that there were months over the past year in which they did not have enough food to meet family needs. On average, they reported not being able to meet family food needs in 12 out of the 30 days preceding the interview. Finally, about 15% of households reported that there were times when a child in the household did not eat the whole day because of the lack of food.

Table 12. Food security

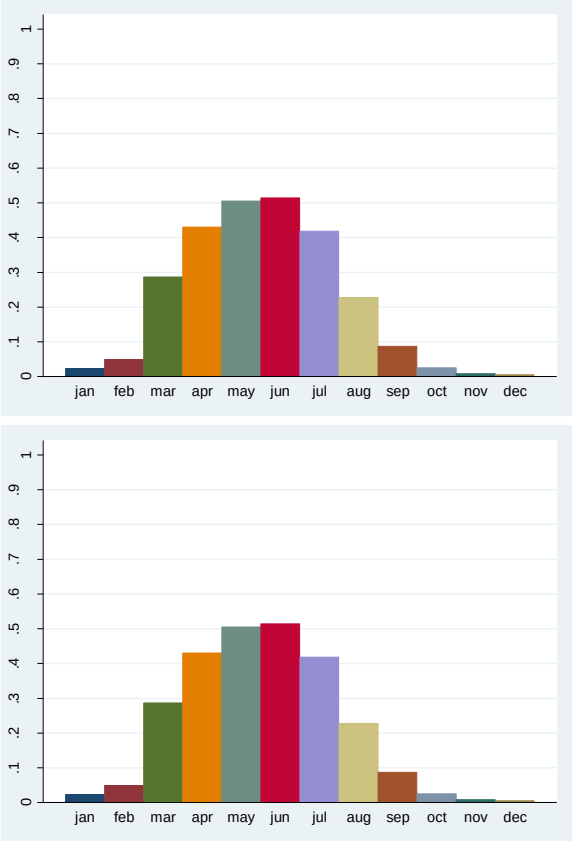
	MV	CV
Not enough food in any month over last year %	82.2	84.6
Days with not enough food over last 30 days	12.2	12.9
Any day a child went hungry the whole day %	16.4	14.8

*Difference statistically significant at 10%, ** significant at 5%, ***significant at 1%

82. There is a strong seasonal pattern in food availability. This is the consequence of a subsistence economy where there is only one cropping season and limited opportunities to save.[29] Households were asked about the months when they did not have enough food to meet their family's needs. Few families complained of food scarcity from September to February. However, March to August and particularly April to July seemed to be the 'hungry season' in which about 50% of households reported not having sufficient food (see Figure 2).

[29] The Reality Check suggested that households had space to store crops. As such, the problem was less about the lack of storage and more about not having enough produce to store. Observations suggest that in the hungry seasons some poor households borrow bags of food from more wealthy households and return them doubled at harvest time.

Figure 2. Percentage of households with not enough food by month of the year



4.5 Agriculture

83. Most households in the study area are primarily farming (90%). Farmers cultivate an average of three hectares of land scattered across three different plots.[30] Farmers are mostly subsistence farmers and the proportion of agricultural produce sold on the market is only about 20%. Farming also represents the largest share of total income (60%) followed by livestock (25%). The next most important source of income is having a microenterprise. About 20% of households have microenterprises, mostly in the areas of trading (40%), retailing (20%), and agricultural processing. Other sources of income are negligible. Less than 3% of households had a member engaged in any paid work over the year preceding the interview, and the share of transfer income, including remittances, over total income is less than 1%.

[30] Further research is needed around the reported estimates of landholding size: It is possible that farm sizes are over-reported by illiterate farmers. Agricultural officers interviewed assess the typical farm of a poor farmer to be in the region of 1-3 acres. See also: USAID, 2009. Ghana: Trade and Investment Program for a Competitive Export Economy (TIPCEE) GIS Work (http://www.cop-horti.net/IMG/pdf/TIPCEE_GIS_Work_Feb_2009.pdf) and Hainmueller, J, M Hiscox and M Tampe, 2011. Sustainable Development for Cocoa Farmers in Ghana, Baseline Survey. MIT and Harvard University (https://www.responsibleagroinvestment.org/sites/responsibleagroinvestment.org/files/Ghana%20Cocoa%20Baseline%20Report_Jan%202011.pdf).

Table 13. Farmers and income sources

	MV	CV
Farmers %	91.0	95.2**
Share of production sold in the market %	21.9	24.6
Farming share %	61.3	61.3
Livestock share %	26.3	23.6

*Difference statistically significant at 10%, ** significant at 5%, ***significant at 1%

84. Agricultural activities are conducted under very risky circumstances. All households report having been affected by a shock of some type during the year preceding the survey. Droughts, floods, and crop failures are particularly common (Table 14). There appears to be a contradiction between the smaller reporting of drought in the MVs at the same time as farmers in MVs are reporting a higher incidence of crop failure. These contrasting observations could be the result of seasonality (Section 5.4). MV households were interviewed in May and June and thus were in large part unable to report the outcome of the current crop as farmers in the CV areas who were interviewed over the months of September and October.

Table 14. Households affected by economic shocks

	MV	CV
Drought %	76.0	83.0*
Floods %	57.2	54.7
Livestock death %	86.5	73.6***
Crop failure %	72.5	63.6**

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

85. Most income is spent and very few households are saving. Only about 15% hold a bank account or are members of a *susu* group[31] and less than 5% of households have taken a out a loan over the 12 months preceding the interview.

Table 15. Savings and loans[32]

	MV	CV
Household has a bank account %	15.6	10.8*
Household is member of *susu* group %	15.0	8.5**
Any loan over last 12 months %	4.9	3.3

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

86. The most common form of savings is animal holdings. More than 80% of households save at least one animal such as chickens, goats, or guinea fowls. The median value of all animal stocks per household is about $250-300 PPP, which is more than half of the average annual per capita expenditure. The RCA indicated that poorer households have very few animals and no large ruminants. They keep smaller ruminants such as goats (no more than three) and several fowl only as easily liquefiable assets or for sacrificial use.

[31] *Susu* literally means "small small" and refers to traditional informal rotating savings and credit schemes.
[32] The phrase "household" here refers to any household member having a bank account, *susu* membership or a loan. Typically, bank accounts and loans are taken out by an individual rather than a household as a whole.

Table 16. Animal stocks

	MV	CV
Households with animals %	85.2	80.4
Average number of cows[33]	3.1	2.2**
Average number of goats	3.8	3.5
Average number of chicken	6.8	7.4
Average number of guinea fowls	3.7	2.7
Median value of animal stock ($PPP)	301	234

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

4.6 Education

87. Attendance[34] of primary school aged children is lower in the area than compared to the rest of Ghana, but not by a large margin. The net attendance rate in Ghana was 74% in 2008 (both in urban and rural areas) compared to 60% observed in the MV sites. The RCA indicated high motivation among parents to send their children to school even though they themselves had not received any education.

88. However, attendance ratios are very low in JHS and Senior High Schools (SHS). The RCA found that whilst primary education was highly valued for all, families chose which children should continue to secondary level based on the child's potential to succeed. The investment required for secondary education, distances to access secondary schools, and the fact that children completing primary education are often 'old for grade' act as disincentives to continue.

Table 17. Summary of education indicators

	MV	CV
% over 5 ever attended school	49.9	46.3
NAR primary	60.5	68.9*
NAR JHS	9.7	15.4**
NAR SHS	5.0	6.8

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

Note: These are net attendance ratios calculated as the proportion of children in school among children in the school-level specific age group (three age groups were used: 6-11, 12-14, and 15-18).

89. Interestingly, there are more girls than boys in school at all grades: 69% against 63% in primary school, 17% against 10% in JHS, and 7% against 5% in SHS. All differences are statistically significant. This contrasts with official enrolment data, which show a larger percentage of boys in school. The difference is explained by the fact that the national census and survey data look at different types of schooling: enrolment in the case of the census and attendance in the case of the survey. The latter is a better indicator of actual school attendance. The RCA study observations in schools and discussions with families and teachers confirmed the higher school attendance of girls. Possible explanations include that boys are more likely to experience punishment at school for poor behaviour, attendance, or study and thus become de-motivated. They are also eager to earn incomes as soon as they

[33] Note that 'cows' does not include bulls that are reported separately in the questionnaire. Observations of the enumeration by the QA team suggest that if anything, farmers are under-reporting the numbers of animals for a variety of reasons. There is however a minority with a higher number of cows that leads to a high sample average.

[34] Attendance is based on household interviews rather than school records. Respondents report whether they attended school at any time in the previous year or month. This is consistent with attendance rates used by DHS, MDGs, and the World Bank, and is considered to be more reliable than other figures.

become physically strong enough to work in order to purchase mobile phones, clothes, and snacks for themselves.

90. Literacy rates among adolescents are also higher among girls compared to boys. These rates are calculated as the percentage of young men and women (15 to 24 years of age) that are able to read a simple English sentence. These rates are much lower than in the rest of Ghana (80.1%) and the gender pattern is different from the rest of Ghana where young women are less likely to be literate (76.8% of girls against 84.0% of boys) and less likely to be in school.

Table 18. Literacy rates among adolescents (15-24)

	MV	CV
All	33.1	34.7
Young men	25.5	29.6
Young women	37.7	38.0

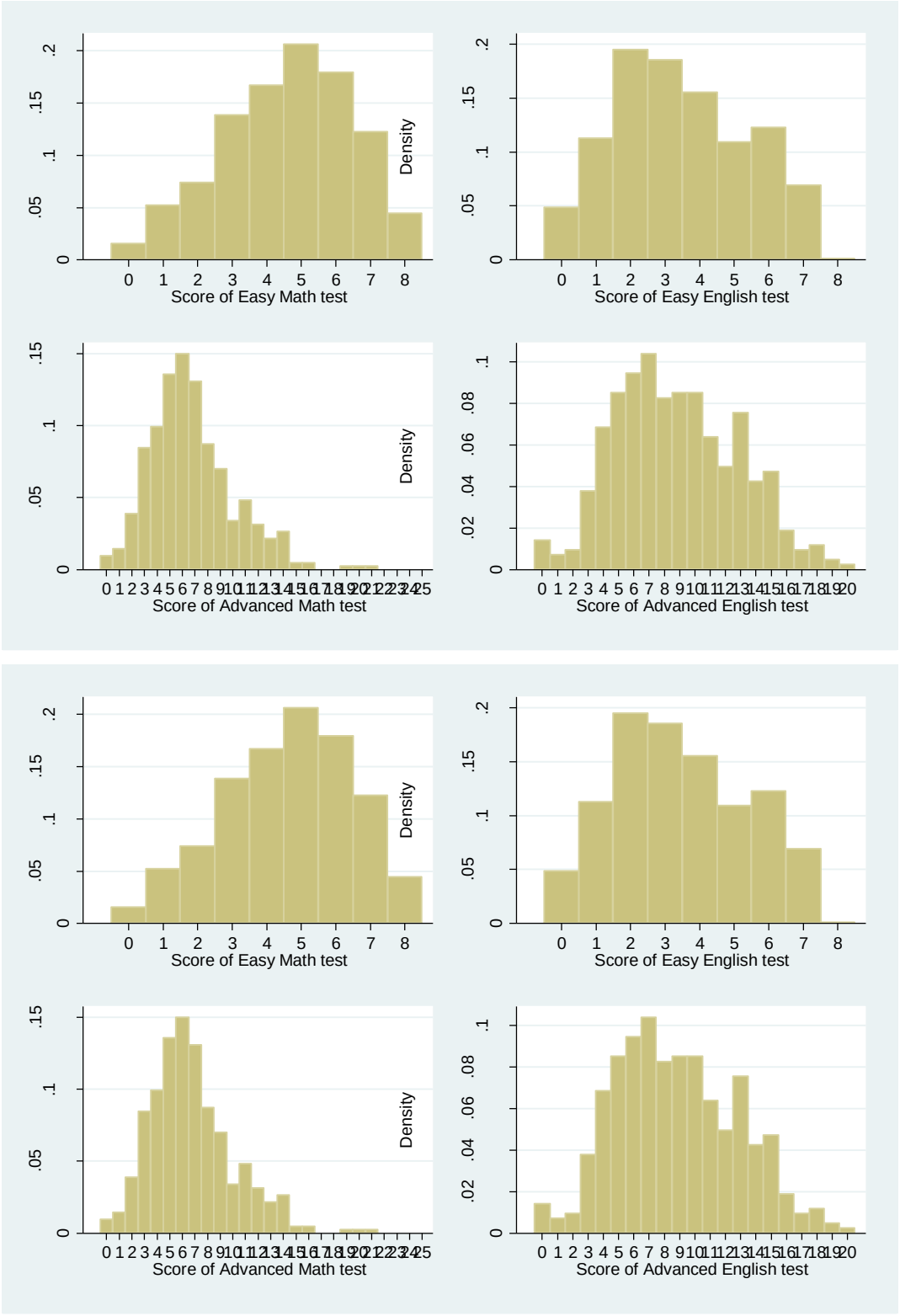
*Difference statistically significant at 10%, ** significant at 5%, ***significant at 1%

Note: These rates are calculated as the proportion of young adults (15 to 24) able to read aloud all the words of the sentence "The child is playing with the ball."

91. It is apparent from the maths and English tests that the quality of schooling is poor. Few children aged 9-19 who have ever attended primary school are able to complete tests based on eight simple arithmetic operations and the understanding of simple English sentences. Similarly, children aged 12-19 who ever attended JHS performed poorly in advanced English and maths tests (Figure 3). The FGDs identify hunger and seasonal streams[35] among the barriers to consistent attendance. In many of the communities with schools, teachers prefer to live in the distant district capitals as they are better served with social amenities, resulting in them being routinely late. In some of the FGDs, participants noted teachers' frequent absenteeism during the rainy season and major football festivals. Where possible, the well-to-do enrol their children in non-state schools to improve their prospects of getting a good education.

[35] This is not the same as the seasonal flooding. These are streams that are not permanent and flow intermittently.

Figure 3. Test scores in the study population



92. The RCA noted teacher shortages as well with high levels of teacher absenteeism in five of the six village primary schools in addition to empty teacher accommodation. There was also a chronic lack of

teaching materials and overcrowded classrooms. Children complained that they did not understand the lessons and 'teachers do not teach properly.'

4.7 Health

93. Mortality rates are measures of child survival and a reflection of the general living conditions of a population. Mortality rates were calculated over the five-year interval preceding the survey using the synthetic cohort probability method used by the Demographic and Health Survey (DHS). Bootstrapped standard errors were calculated for the estimates and the results are presented in Table 19. Mortality rates in the study area are high by international standards and higher than in the rest of Ghana. There are large differences in the mortality rates observed in project and control areas. Mortality rates are not measured with precision because child deaths are rare events and the standard errors of the estimates are large. Precision in the estimates depends on the size of the sample; while the DHS estimates are based on a sample of 2,992 children the MV area contains 1,367 observations and the CV area contains 3,033. It is possible therefore that the small sample size in the MV area is underestimating the true mortality rates. An alternative explanation is that maternal and child health is better in the MV areas but this is not fully supported by other health indicators collected by the survey.

Table 19. Mortality rates (per 1,000 live births)

	Pooled data	MV	CV[36]	P-value
Neonatal	35.6	28.1	39.5	0.090*
	(0.004)	(0.007)	(0.005)	
Post-neonatal	14.5	22.1	10.6	0.011**
	(0.004)	(0.007)	(0.005)	
Infant	61.1	43.2	69.9	0.005**
	(0.005)	(0.008)	(0.007)	
Child	36.0	28.7	37.8	0.144
	(0.004)	(0.007)	(0.005)	
Under 5	91.9	70.6	105.1	0.004**
	(0.006)	(0.010)	(0.008)	

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

Note: Mortality rates were calculated using the synthetic cohort probability method. Standard errors and p-values are obtained by bootstrapping and resampling the variance to take into account the two-stage cluster design of the datasets. Stata do-files for the calculation of mortality rates, standard errors, and p-values are available on request.

94. The RCA revealed that families regard themselves as 'blessed with good health' and use a mixture of traditional practices and conventional medicines that are available on the market to self-treat. They resort to government health provision only when illnesses are considered critical and tend to circumvent local health clinics by going directly to better resourced district facilities. Relatively few have health insurance, indicating that they do not need it either because they are healthy, have more accessible alternatives (e.g. local herbalists, soothsayers, medicine sellers) or that they consider the cost (both actual and opportunity cost of getting the card or of accessing distant health facilities) prohibitive. However, some families do have insurance for their children.

[36] The CV figures are higher than the MV and consistently higher than the Ghana figures. The explanation for this is not unclear.

95. Both qualitative modules also indicated a strong preference for using traditional birth attendants (TBAs), many of whom have received some formal training in the last 10 years. The PRA study suggested that TBAs are among the most respected people in the village. Discussions with families and TBAs suggest that families' preference is based on trust, their willingness to assist any time of day or night, the preference for giving birth among family at home, and the TBA's familiarity with the family. It was indicated that they refer difficult cases to health centres and are quick to seek assistance using mobile phones when they run into unexpected problems. Few incidences of neonatal or postneonatal deaths were noted by families or TBAs.

96. Undernutrition rates are high but strikingly similar to malnutrition rates observed for the whole of Ghana by the 2008 DHS. The 2008 DHS data report higher rates of stunting and underweight for the northern regions which may be reconciled to our data (from 2012) by considering the declining trends in undernutrition in the area.

97. The RCA noted that very few mothers feed their babies colostrum even though the 'nurses tell them to' as the belief that it is 'dirty' still prevails. They rarely practise exclusive breastfeeding and provide drinking water from when babies are only a few days old.

Table 20. Scores across MV and CV areas

	MV	CV
Height-for-age Z-score	-1.22	-1.18
Moderate malnutrition <-2	27.0	28.1
Severe malnutrition <-3	7.0	12.8**
Weight-for-age Z-score	-0.89	-0.85
Moderate malnutrition <-2	14.0	16.2
Severe malnutrition <-3	3.0	5.2**
Weight-for-height Z-score	-0.29	-0.23
Moderate malnutrition <-2	4.0	5.1
Severe malnutrition <-3	0.0	1.0**

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

Note: Z-scores and rates were calculated for children from 6 to 59 months of age. Z-scores were calculated using the reference population of the new WHO sample. There are differences in standard deviation units from the reference population norm.

98. Only about one third of children are regularly included in state-run preventative health services. Indications of the reach of health services is the proportion of households that reported being visited by government health workers and the proportion of children who reportedly received supplementary vitamin A and deworming tablets (Table 21). However, the RCA noted that the latter are school-based programmes that parents are not necessarily knowledgeable about.

Table 21. Coverage of state health services

	MV	CV
% taking vitamin A last six months	61.6	63.8
% taking deworming treatment last six months	33.6	32.6
% visited for family planning	24.8	25.4
% visited by health visitor for general care	35.3	39.7

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

99. The RCA found that interest in family planning was very low and there are strong social norms endorsed by both men and women to produce many children (Box 2).

Box 2. Reluctance to adopt family planning

Health workers came to one of the RCA villages two weeks before the study to speak about family planning to the men of the village, but the men refused to listen and sent them away ('God wants them to give birth to as many as they can'). Women explained that if a mother does not keep producing babies, the neighbours jibe and tease, suggesting the husband is impotent or that she is barren. However, once 'the eldest son is married then the husband and wife stop sleeping together – to have a baby then is not right and people will call you names' (older woman). In the Buili communities, there was also a sense that 'God will decide the number and we will take' (woman). Besides, 'you can die at any time, better to give birth to as many children as possible first' (woman). Several mothers told us they 'enjoy giving birth to many children.' Furthermore, as the Fulani men explained, 'children are a gift of God, you might block an important person coming into the world,' (men).

100. Anaemia rates are very high but are comparable to those in the rest of the country, which are also similar to or lower than other West African countries (Table 22). Following DHS standards, mild anaemia is calculated as the ratio of children with haemoglobin below 11 g/dL, moderate anaemia is haemoglobin below 10 g/dL, and severe anaemia is haemoglobin below 7 g/dL.

Table 22. Prevalence of anaemia among children under five

	MV	CV	Ghana[a]
Mild anaemia	74.2	84.0**	77.9
Moderate anaemia	45.7	60.0**	55.0
Severe anaemia	3.7	5.2	7.4

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

[a] Data from the DHS 2008.

101. Anaemia may in part be related to malaria prevalence. Despite the high rate of households owning mosquito nets (above 80%), the incidence of malaria is rather high (Table 23).

Table 23. Mosquito nets and incidence of malaria

	MV	CV
[a] Household has a mosquito net %	81.3	90.2***
Malaria incidence among children under 5	22.3	23.4

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

[a] Note: this includes all bed nets whether they are insecticide treated or not.

102. The RCA also noted a high level of mosquito net ownership, in some cases there were more nets than necessary, but less than 25% of families used them due to sleeping outside. Families recorded using nets in the wet season and when mosquitoes disturbed their sleep. A connection was not made between a reduction in malaria incidence and using bed nets.

103. Most of the population has access to improved drinking water, but only 10% uses an improved toilet facility. Even though boreholes and wells are reasonably accessible, both the RCA and the FGDs noted that poorly maintained facilities result in time-consuming queues or further distances to collect water from working facilities. In some communities, there was a strong preference for well water over borehole water because people claimed the latter tasted bad. All of the RCA study families practised open defecation and the RCA team observed very little use of the few improved toilet facilities.

Table 24. Access to improved water and sanitation

	MV	CV
Households with improved water %	73.2	72.2
Improved sanitation facility %	10.1	10.4

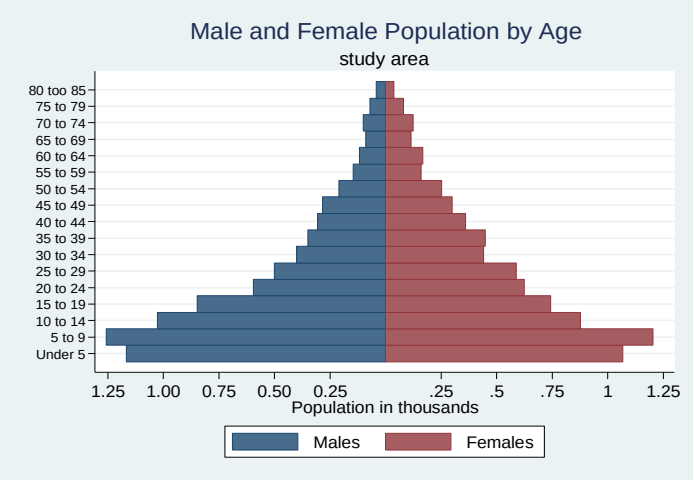
*Difference statistically significant at 10%, ** significant at 5%, ***significant at 1%

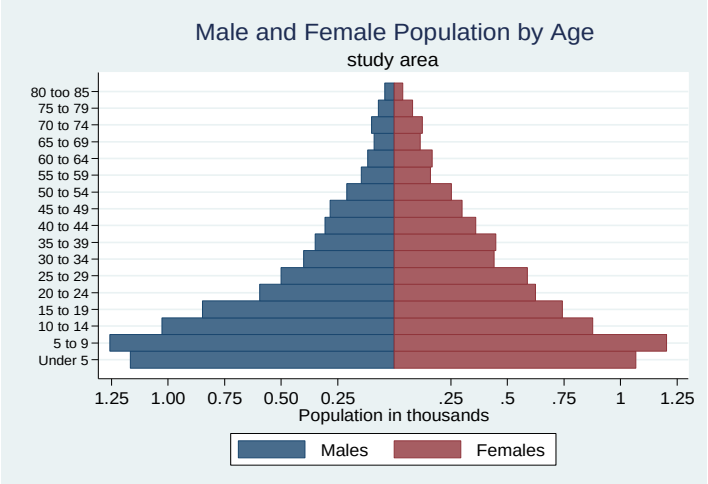
4.8 Gender

104. The state of maternal health is very poor in the MV area with only 27% of births attended by a skilled professional, 66.7% of women visiting a health facility before delivery, and only 10% using contraceptive methods (Table 6). These values are very low not just in absolute terms but also in comparison with the rest of the country. The data do not show that girls are disadvantaged in terms of access to education. On the contrary, a larger proportion of girls attend school compared to boys at all levels of education.

105. The demographic structure of the population by gender shows interesting patterns (Figure 4). There are more boys than girls up to the age of 20, after which there are more women than men at all ages. It is difficult to explain this without further data analysis, though it may relate to different patterns of mortality and migration between men and women.

Figure 4. Population pyramid in the study area





4.9 Social networks

106. There are extensive social networks in the area with about 80% of households reporting that important people (e.g. relatives [55.2%] or friends [21.2%]) live in another village. In 50% of cases, help (sought or provided) consists of general advice or farming advice, while in 30% of cases it consists of gifts, and only in a few cases it consists of borrowing or lending. This information will become relevant and useful to estimate the size of spillover effects from the MV area to the neighbouring areas.

Table 25. Social networks

	MV	CV
Any important people living elsewhere? %	76.0	83.5**
Of which distant relatives %	55.2	64.2*
Of which friends %	21.2	20.9
Asked any help over last 12 months? %	45.0	45.7
Provided any help over last 12 months? %	53.0	50.4

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

107. Both the RCA and FGDs noted that the traditional reciprocal arrangements for farm work and house construction during the non-farming season still operate, but that the need for cash is gradually eroding this system. Some households indicated that they have to pay wages whereas before they would have exchanged labour and food or part of the crop as payment.

108. More than half of the RCA families and their neighbours own mobile phones, which are considered the primary way of maintaining social networks and sharing both advice and information.

4.10 Expectations and time preferences

109. Attitudes towards risk, such as expectations of survival, income, education, and educational returns, affect behaviour and choices. It is believed that many life decisions such as educational or occupational choices or even the purchase of fertiliser (Duflo, Kremer, and Robinson 2011) are influenced by 'impatience.' Impatience can be defined as people's inability to postpone current pleasure in exchange for future benefits. Rational and patient people will buy health and travel insurance; they will save for their studies or for old age and will forgo other current pleasures to obtain future benefits. Impatient people on the other hand are not able to exercise self-control and end up uneducated and without insurance. Extremely poor households tend to be highly 'impatient' because they are deprived of almost everything and are forced to spend the few means they have, an idea that goes back to Fisher (1930). The evaluation estimated time preferences among a sample of approximately 1,000 commercial farmers using hypothetical lotteries.[37] Famers were given hypothetical choices to make between present and future rewards in order to assess their ability to postpone current gains for future highest benefits. It was found that about 30% of farmers apply a zero discount rate whilst the average monthly discount rate is not far from discount rates observed in other contexts by similar exercises. At first sight these farmers do not appear to be particularly impatient. The discount rate decreases with the time horizon, thus pointing to hyperbolic discounting. Hyperbolic discounting is commonly observed in behavioural experiments and reflects an inconsistency in people's evaluation of future rewards. People tend to discount future rewards

[37] Hypothetical lotteries use the matching task method (rather than real rewards) to test people's choice preferences between immediate and (higher) delayed rewards.

more heavily than current ones. In other words, they valuate more highly rewards coming sooner rather than later. Farmers in our sample appear to conform to this type of behaviour.

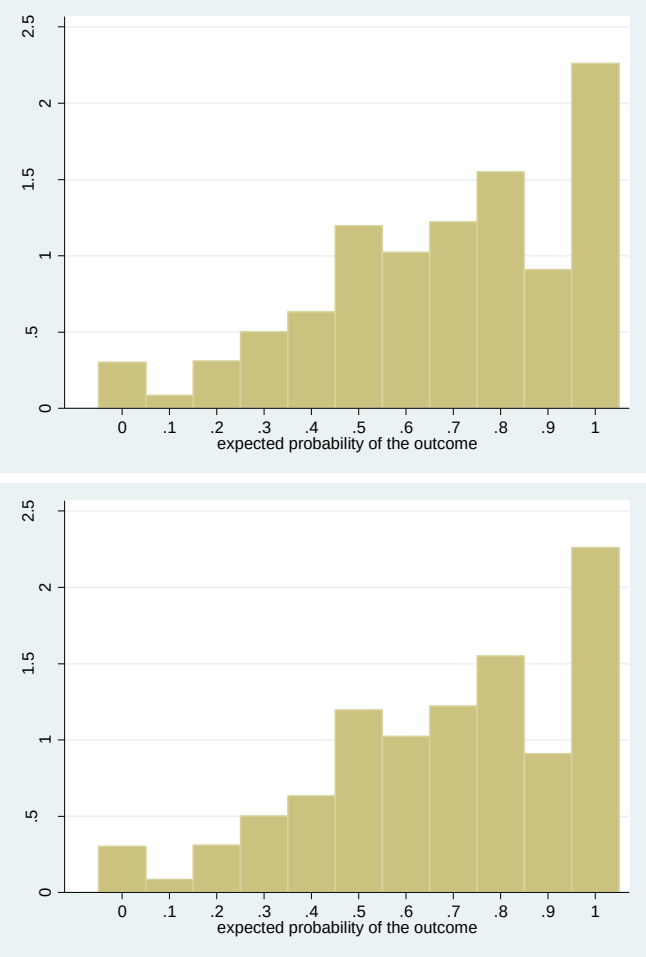
Table 26. Discount rates and 0 discount rates for the whole sample

	% Monthly discount rate is zero	% Monthly discount rate
1 month horizon	0.33	0.087
3 month horizon	0.14	0.075
6 month horizon	0.07	0.071
12 month horizon	0.06	0.055

*Difference statistically significant at 10%, ** significant at 5%, ***significant at 1%

110. In order to make the lotteries closer to a real life situation, the hypothetical lotteries for the time preference exercise were based on expected agricultural output. Farmers were asked first to indicate the range of possible agricultural output under the best and worst scenarios. They were then asked to rate the chances of obtaining the midpoint of this range. The expected probability for the average agricultural output is displayed in Figure 5. Farmers show different degrees of uncertainty regarding their expected production. Uncertainty plays a significant role in farmers' production decisions and the variability in uncertainty among farmers will be both explained in the analysis and used to explain production choices.

Figure 5. Expected probability of the average agricultural output



111. The evaluation also calculated parents' and children's wage expectations. Parents, and their children separately, were asked to rate the current average daily wage in the area and in Accra for different levels of education. Standard cost-benefit analysis of educational choices in Becker's tradition posits that parents compare the cost of schooling to the expected future income stream from schooling in order to decide about schooling their children. How expectations about future incomes are formed is not known but there is a consensus that they should play an important part in schooling decisions. Parents expected wages are reported in Table 27. The wage expectations are not far from the actual wages observed in the field. The community questionnaire finds an average male wage of five cedis per day for agricultural labour.[38] In expectation terms, the wage increases with educational level and is believed to be more than twice in Accra compared to the study area. It is puzzling that control areas have much higher wage expectations than the MV area, which cannot be explained by seasonality because these interviews were conducted at the same time of the year in the two areas.

Table 27. Wage expectations in Ghanaian cedis (parents)

	MV	CV
Daily wage primary education	4.94	6.91***
Daily wage secondary education	6.72	13.13***
Daily wage primary education Accra	12.26	18.80***
Daily wage secondary education Accra	18.19	30.49***

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

38 This appears low, and is around 2.5 dollars per day. It is however higher than some comparable areas, such as rural wages in dry land India (less than $1 per day for men and 50 cents for women).

5. Characteristics of the data

5.1 Balancing tests

112. The success of the DD strategy adopted by this study rests on the validity of the assumption that project and control villages are similar. The selection of the control communities was conducted by matching control communities to project communities using village-level variables. This selection is clearly not as good as a random selection and it needs to be assessed whether there are any large and statistically significant differences between project and control villages. In principle, DD analysis is only based on the assumption of parallel trends. Baseline equality in the levels is not needed. What really matters for the identification of the project effect is that the rate of change over time in the variables of interest is the same in the two groups. In practice, for many variables trends and levels are related in such a way that a specific rate of change (slope) of an outcome is normally associated with a specific level of a determinant. Therefore, a comparison of project and control communities is conducted for both in the levels and in the trends.

113. A list of variables was selected at the locality and household levels in order to test differences between different samples. The surveys collected data on several dimensions and a full balancing test would require hundreds of variables. It was instead decided to select a representative set of variables and to do so blindly as much as possible. A large pool of candidate variables was selected first that are known to be relevant to the analysis of the results because: (i) They are key outcome variables, such as poverty; (ii) They are determinants of the outcomes that will be used as control variables when estimating treatment effects, such as land; and (iii) They are variables determining participation in project activities that could be used to match households or individuals at the analysis stage, such as demographic structure. At a second stage the most relevant variables were retained, such as poverty and household size, and blindly selected one or two variables within a group of similar variables. For example, 'fetching water' and 'collecting wood' were blindly taken from a pool of six 'time use' variables. The final list of tested variables includes:

- Final and intermediate outcomes
- Factors determining the outcomes
- Factors affecting project participation

114. In conducting the comparisons, both the size of the difference and its statistical significance will be looked at. Absolute values and percentage differences will be used when these are easily interpretable. When interpretation of the absolute values is difficult, for example in the case of test scores, standardisations of the means by the standard deviations will be used. The convention of considering a difference of 0.2 standard deviations between groups as 'small' and a difference of 0.5 standard deviations as 'medium' (Cohen 1988) is adopted. A difference of 0.1 is considered 'very small' and anything below this difference irrelevant. With populations normally distributed, with equal variability and equal size, these differences can be interpreted in terms of percentage shift of the normal distribution curves between the two groups. For example, a standardised difference of zero means a perfect overlap between the bell curves of the two populations; a difference of 0.2 is equivalent to a 15% shift and a difference of 0.5 is equivalent to a 33% shift.

115. The sample for the study was selected through a two-step design in which communities were first selected and then households were randomly selected within communities proportional to the size of

the communities. This self-weighted design, in which large communities are represented by a larger number of households, does not require the use of weights at the analysis stage. The cluster structure of the sample however generates an artificial reduction of the variance though the intra-cluster correlation of the variables. The cluster structure of the data is taken into account by calculating standard errors that are adjusted by the values of the intra-cluster correlation coefficients. The fact that the matching procedure was conducted within each district (separately in Builsa and West Mamprusi) by controlling for regional differences is also taken into account. In this way the standard errors are reduced and differences are more accurately estimated. In practice, ordinary least squares (OLS) regressions were run for each variable including a dummy for the treatment status, a dummy for the region, and a correction for the cluster structure of the standard errors.[39]

116. This exercise was conducted using (i) Village-level variables collected by the community questionnaire and (ii) Community averages of household-level variables collected through the household and the adult questionnaires, anthropometric, and educational modules. These test differences are used to evaluate the accuracy of the matching performed at the community level in order to select control sites. In addition, the household level differences across the entire samples are tested. Communities have different sizes and were selected proportionally to their population size. This latter comparison better captures the differences between the study populations. These results are presented in Tables B1, B2, and B3 in Appendix B. In the tables four differences are presented with respective P-values. Stars have the usual interpretation of increasing statistical significance (one star at 10%, two stars at 5% and three stars at 1%). The differences are as follows:

- CV-MV: the difference between the control group and the project group
- CVN-MV: the difference between the nearby control communities and the project group
- CVF-MV: the difference between the faraway control communities and the project communities
- CVF-CVN: the difference between the faraway and nearby control communities

117. If the data had been collected from randomly selected samples of communities, as in a cluster randomised trial, it would be possible to accept up to 10% of statistically significant difference below the 10% threshold, because this is the number of differences that could be obtained just by chance. A percentage of variables statistically different at 10% in excess of 10% of the total number of variables considered is a sign of structural differences across the samples. The percentage of statistically significant (at 10%) differences is reported in Table 28 for each of the Tables B1, B2, and B3 in the Appendix.

[39] The stata code for, for example, poverty is: *regress poverty CV Builsa, cluster (community)*. In this example, the reported differences and p-values are those associated with the CV coefficient.

Table 28. Percentage of significantly different variables at 10%

	CV-MV	CVN-MV	CVF-MV	CVF-CVN
Table B1 Community level variables	30%	26%	20%	23%
Table B2 Community level household variables	23%	21%	25%	17%
Table B3 Household variables at household level	23%	23%	25%	17%

118. Mean differences were tested between project and control villages using 30 variables selected from the criteria outlined above. Nine statistically significant differences were found at the 10% level between the project and control communities corresponding to 30% of selected variables. Notable differences include cropping patterns (in the MV maize production is predominant against millet in the control group); social organisations (there are more farmer cooperatives in the MV area but fewer women's groups); and population covered by health insurance (this is higher in the control areas).

119. The exercise was repeated using community-level averages of household data collected by the household questionnaire and other modules using 53 variables. Community-level variables (such as the availability of a primary school) and community averages of household level data (such as test scores) do not need to be strongly correlated and therefore a good (or poor) matching on the community-level variables does not necessarily imply a good matching on the household level variables. 12 statistically significant differences were found between the samples of MV and CV communities corresponding to 23% of the selected variables. Observed differences in CV areas include: a larger number of in-migrants, higher primary school attendance, longer time spent in household chores, poorer credit access and larger share of farmers, larger social networks, higher use of mosquito nets, higher incidence of anaemia, better wage expectations, and proficiency in English tests.

120. A test using the full household data was conducted using the same process. The results are very similar to those obtained in the previous exercise. 53 variables with 12 statistically significant differences were found between MV and CV areas, representing 23% of all variables. This percentage is larger than the 10% that was expected to be found through chance. There are several possible explanations for why this occurred. First, some differences are found among seasonally-sensitive variables, such as episodes of diarrhoea in the last two weeks or anaemia incidence and are probably the result of the different timing of the surveys in the project and control areas (see Section 5.4 for a discussion of seasonal issues). Second, some could be a reflection of differences in the socio-economic characteristics of the two areas. For example, the large difference in expected wages consistently held by parents and children that is not affected by seasonal bias. Finally, some differences are simply due to chance.

121. Are the nearby CVs more similar to the MVs than the faraway CVs? By simply counting the number of statistically significant differences, faraway communities are not more different than nearby communities as the number of statistically significant differences is very similar for the two groups.

Are the nearby and faraway villages more similar to each other than to the MV villages? Only slightly so as there are fewer statistically significant differences between faraway and nearby communities. Households from faraway communities are different in the following ways: they are less food secure and display lower weight-for-age Z-scores, they have much larger social networks, they have a larger average number of years of schooling, and they show higher use of mosquito nets and visit health facilities more frequently (a result mirrored by the higher access to health facilities emerging from the community-level data). Note that no differences were found between faraway and nearby communities in those variables that were believed to be affected by seasonal bias such as shocks, time use, incidence of anaemia, and diarrhoea.

5.2 Balancing by matching

122. In the previous analysis of village and household characteristics it was found that there are differences between the MV and CV groups in the averages of the project outcomes and in the averages of the determinants of the outcomes. These differences amount to 30% of characteristics in the case of village-level variables and to 23% in the case of household-level variables. The samples of villages and households in the MVs and CVs are unbalanced. This poses two questions regarding the validity of the data collected. First, are the differences observed large and can they be safely ignored? Second, can the differences be balanced using matching methods?

123. With respect to the first question, the per cent differences reported in Section 5.1 should be interpreted as in excess of 10% differences that would have been found anyway because of chance error. The number of observed differences, particularly in the case of the household-level variables, is therefore not too large, though clearly larger than the number that would have been obtained by randomly allocating the intervention within a large pool of candidate villages. As for the size of the differences, it should be noted that these mostly amount to few percentage points for binary variables, or less than 0.2 standard deviations for continuous variables. These differences are therefore 'small' and do not suggest that there are large structural differences between the project and the control groups. There are however some 'large' differences, namely in mortality rates (2-3% lower in the MVs), school attendance (6-8% lower in MVs), ownership of mosquito nets (9% lower in MVs), and haemoglobin levels (0.3 standard deviations higher in the MVs). It is difficult, without further data analysis, to tell whether they are the result of chance, seasonality, or structural differences between the areas.

124. The second question was addressed by running some matching experiments using the baseline data. Successful matching of observations in impact analysis rests on the ability to identify the main observable determinants of participation into a specific intervention. The first stage of analysis consists of developing a behavioural model that explains participation in a project activity using a series of 'determinants of participation.' The list of these determinants will vary depending on the level at which the outcome is observed. For example, different factors will explain the selection of villages and households. Secondly, the determinants should be able to capture the targeting rules and self-selection processes. For example, different determinants will be employed for explaining child participation in the supplementation of vitamin A and for farmers joining cooperatives promoted by the MVP. Finally, the determinants must not be influenced by project operations. If this happens, matching fails as the matched observations are similar to the project observations only after the intervention. The MVP consists of a large number of activities targeted to different groups: households, children under five, pregnant mothers, maize farmers, school children, and so forth. In this experiment we decided to perform matching at the most general level. The household as a unit

was used for observation and the ultimate goal of poverty reduction was considered in setting up the participation model so as to exclude any targeting and self-selection rules.

125. At the analysis stage matching will be conducted employing a richer set of variables depending on the specific outcomes investigated. For example, in the estimation of nutritional impacts the characteristics of the child, such as age and gender, will be included in the participation model but they are not included in the more general model below looking at household-level poverty. A probit selection equation was conducted at the household level using household and village-level variables. Some variables are statistically significant pointing to differences between the MV and CV samples that were observed in Section 5.1. Households in MVs are preponderantly of Builsa ethnic group, are more likely to be headed by a female, are less likely to be farming, and are less likely to be food insecure. They are more likely to rely on social networks, but are also more likely to be member of *susu* groups. MVs are more likely to lack a JHS and a market. Male agricultural wages tend to be lower in MVs and the areas are less visited by government agricultural extensionists. Finally, MVs are more likely to have an NGO operating in the area. The picture emerging from this model is that there are obvious differences between MVs and CVs that suggest MVs are worse off compared to CVs in some respects.

Table 29. Probit selection equation for household participation in MV

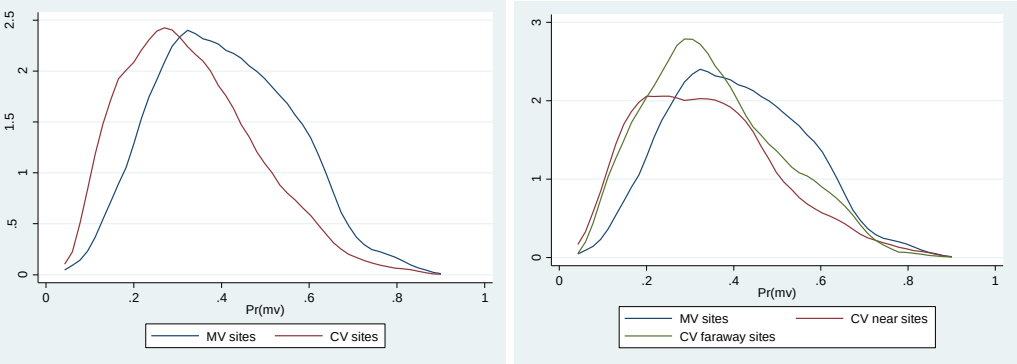
Variable	Coefficient	P-value
Household size	0.004	0.720
Number of children	-0.004	0.906
Female headed household	-0.260**	0.010
Polygamous household	0.128	0.131
Age of head of household	0.001	0.674
Builsa ethnicity	0.409**	0.017
Mampruli ethnicity	-0.454**	0.005
Education of head of household (years)	-0.005	0.647
Value of household assets	<0.000	0.111
Value of animal stock	<0.000	0.685
Farmer head of household	-0.218**	0.039
Household owns a microenterprise	0.016	0.839
Land owned	0.002	0.439
Household food insecure (self-reported)	-0.232**	0.006
Improved drinking water	0.074	0.276
Improved sanitation facility	0.014	0.891
Finished walls	-0.097	0.224
Finished floors	0.129	0.105
Finished roof	-0.006	0.934
Solid social networks	-0.307***	0.000
Member of *susu* group	0.304***	0.001
Affected by drought	-0.131*	0.081
Affected by flood	0.035	0.561
Primary school in the locality	0.275	0.104
JHS in the locality	-0.420***	0.000
Market in the locality	-0.158*	0.061
Distance to market	-0.056***	0.000
Male wage in the locality	-0.126***	0.000
Agricultural extensionist in the locality	-0.174**	0.013
NGO in the locality	0.402***	0.000
Constant	1.072	0.001
Pseudo R-square		0.107
Observations		2173

126. The predicted values of the probit regression (the propensity score) were used to match project households to control households using a kernel algorithm (using the *psmatch2* command in stata); the matching was very successful. A t-test after matching over the determinants in Table 30 shows that matching removed all statistically significant differences between the two samples with the exception of the presence of a market in the locality, which is still significantly different at 0.070 after matching. The average absolute bias is reduced from 11.3 before matching to 3.5 after matching, and

the same probit regression as in Table 30 on the matched samples returns a Pseudo R2 of 0.008 (P-value=0.981) from a value of 0.107 (P-value=0.000).

127. No observation in either sample is extraordinarily different and not comparable to the observations in the other sample. Based on the propensity scores obtained in Table 29, no observation falls outside the region of common support (see Figure 6). This suggests that no observation has to be removed from the data before comparing the two samples to analyse differences.

Figure 6. Distribution of propensity scores in MV and CV sites



128. Finally, matched samples were used to evaluate differences in outcome indicators before and after matching. Three general household-level indicators were selected: poverty, the average number of out-migrants per households, and the availability of mosquito nets in the home. Other outcome indicators, such as children's test scores or anthropometrics, would require different selection models and cannot be compared after matching over the propensity score calculated in Table 30. Balancing on matched characteristics reduces the differences in poverty rates, slightly reduces the difference in the number of out-migrants, and has no effect on the difference in reported availability of mosquito nets.

Table 30. Differences in outcomes before and after matching

Outcome		MV	CV	Difference	St. error	t-test
Poverty headcount						
	Unmatched	0.740	0.719	0.021	0.020	1.04
	Matched	0.740	0.750	-0.009	0.022	-0.41
Out-migrants						
	Unmatched	0.458	0.302	0.156	0.039	3.98
	Matched	0.458	0.334	0.124	0.044	2.81
Mosquito nets						
	Unmatched	0.814	0.902	-0.089	0.015	-5.86
	Matched	0.814	0.904	-0.090	0.018	-5.10

129. In conclusion, there are few differences in observed characteristics between the MV and CV groups. Matching removes these differences, at least at the most aggregate level when comparisons are performed across all households and when there is no targeting or self-selection. All observations are in the region of common support and there is no need to remove observations from the data when conducting the analysis. Yet, it must be recognised that matching was not able to substantially remove baseline differences for two of the variables used in this preliminary experiment: the number

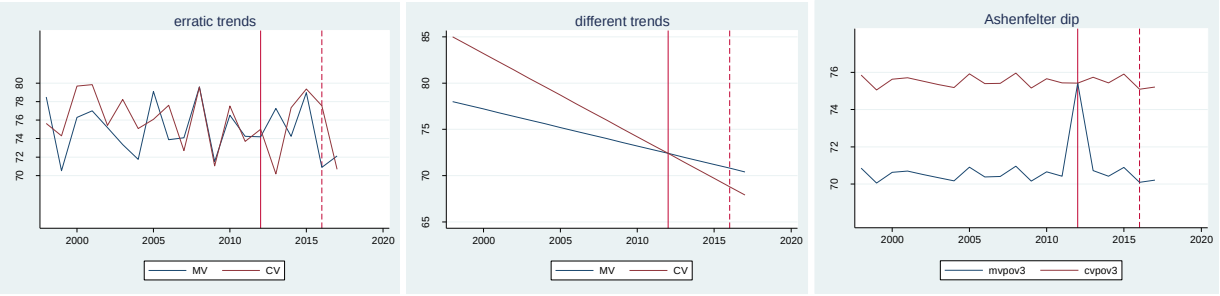
of out-migrants in the household and the availability of mosquito nets. This suggests that other relevant qualifying characteristics of the MVs should be included in the selection model. Inevitably some of these characteristics are not observable, namely seasonality (there is no temporal overlap between the MV and CV samples). Seasonality issues will be discussed in Section 5.4, whilst the validity of the DD estimator with a potentially unbalanced sample because of what is unobservable will be discussed in the next section.

5.3 Difference-in-differences and trend analysis

130. The evaluation design is based on a DD approach comparing the changes in the outcomes in the project villages over time to the same changes observed in a sample of control villages. The validity of this approach rests on the similarity between the project and control observations. The approach is valid if the changes in the outcomes observed in the control villages offer a good description of what would have happened in the MVs without the project. If the outcomes behave erratically in the absence of the programme or if there are strong and different trends in operation in the project and control areas, then DD analysis is no longer valid. In the presence of erratic behaviour of the outcomes or of different trends in the MVs and CVs, DD may find an impact when there is none as well as not finding an impact when there is one. Moreover, the selection of the project sites may be due to more specific reasons, for example because the selected areas had been affected by a drought in the previous year, so that the following natural recovery of the target outcomes is erroneously attributed to the project.

131. A comparison of outcomes and determinants of the outcomes at the baseline may help identify the presence of selection bias. If outcome indicators and their determinants are very different at the baseline then there is a risk that project and control areas are structurally different and that outcome indicators follow different patterns over time. Our analysis of the differences between project and control villages has found that differences are few and small. There are however some differences in outcomes and determinants of the outcomes in the health and education domains, which suggest the presence of structural differences determining different patterns. It should also be observed that balanced project and control samples at the baseline are not guarantees that there is no selection bias. First, differences may exist in the unobservable determinants of the outcomes that will not be removed by DD or matching. Second, the three threats to the validity of the DD estimators outlined above: erratic outcomes, different trends, and Ashenfelter's dip, may occur even when project and control observations are balanced at the baseline. The charts in Figure 7 illustrate this.

Figure 7. Potential DD threats to validity



132. The first chart on the left illustrates the case where poverty rates in project and control villages follow an erratic behaviour because households are affected, for example, by different covariate shocks. Poverty rates happen to be equal in project and control areas at the baseline (solid vertical

line in year 2012) but the DD difference observed at the follow-up (dashed vertical line) would be misleading. The chart in the middle of Figure 7 shows the case in which project and control areas have different trends in poverty reduction. The lines happen to cross at the time of the baseline but the DD estimator based on the follow-up survey would be again very misleading. In this specific case, the project would be shown to have no impact (or a negative impact) when in fact the two areas are not comparable because they are structurally different despite their similarity at the baseline. The chart to the right of Figure 7 shows the Ashenfelter's dip case. For this, suppose that the project areas are definitively better off. Nevertheless, the government selected these particular areas because they had been affected, for example, by a severe drought the year before the intervention. Again the DD estimator would be misleading. It would find a large programme impact whilst in fact the programme has no effect and the project communities are simply naturally reaching at the follow-up their normal status.

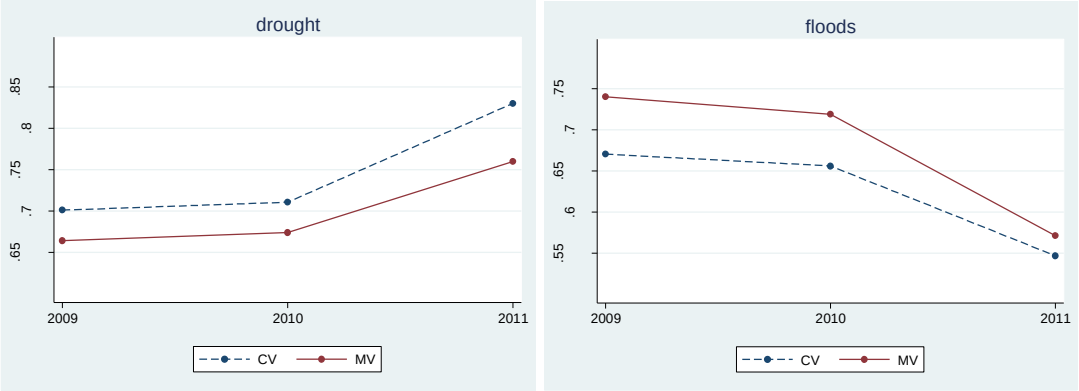
133. Data relating to the years before the baseline are rarely available. In order to build a tentative dataset of trends in the determinants of the outcomes before the baseline retrospective information was collected on household-level sources of incomes (wages, agricultural production, business incomes), animal and land holdings, and covariate shocks. In addition, retrospective baselines can be built for mortality and education outcomes by using birth histories and highest grades achieved that were reported in the questionnaires. Additional village-level trend information on prices and covariate shock was collected in the village questionnaire. Table 31 illustrates the trend variables that can be analysed.

Table 31. Trend data collected by the household and community surveys

Variable	Recall period
Wage income	Each of previous 2 years
Microenterprise profits	Each of previous 2 years
Agricultural output and sales	Each of previous 2 years
Animal holdings	Each of previous 2 years
Land holdings	Each of previous 2 years
Prices (land, animals, fertiliser, labour, and consumer goods)	Previous 2 years and 5 years
Covariate shocks	Each of previous 2 years
Mortality rates by cohorts	Previous 15 years
Education attainment by cohorts	Previous 15 years

134. Trends of most household-level and village-level variables in Table 31 were inspected but evidence could not be found of the erratic behaviour depicted in the left chart of Figure 7, so that this potential source of bias can probably be excluded. Similarly, village-level prices and household-level indicators were inspected in the year before the survey and there were no signs that the MVs were affected by negative shocks in the period before the selection of the intervention areas. The charts in Figure 8 show the incidence of droughts and floods in the MV and CV areas three years before the baseline. In an economy based on rainfed subsistence agriculture, covariate shocks of this type are the main cause of fluctuations in output and related variables. It appears that the MVs are more likely to be affected by floods whilst CVs are more likely to be affected by droughts. However, there is no sign that MVs were more severely affected by any shock before the survey and, more importantly, though the two areas are affected with different intensity, the trends are similar suggesting that the two areas are subjected to the same covariate shocks though in slightly different severity.

Figure 8. Incidence of droughts and floods in MV and CV areas three years before the baseline



135. The occurrence of the bias depicted in the middle of Figure 7 is more problematic and more likely to occur. There is a possibility that the MV and CV areas show different trends in some of the outcome variables. The comparative analysis of the baseline data suggests that education and health outcomes in particular might be differently determined in the two areas.

136. Different trends could not be found in the MV and CV areas in the evolution of household-level incomes and village-level prices. Table 32 shows regressions of income variables on time and time interacted with a dummy for the MV areas. As usual, the regression includes a dummy variable for the Builsa districts and the standard errors (in parentheses) are adjusted for cluster effects. The coefficient for time shows that nominal wage, enterprise income, and the total value of livestock assets increased over the period. The goal of the regression is to test whether this growth was different in MV areas compared to CV areas, which is tested by a statistically significant coefficient for the interaction of MV and time. Trends show that the increase in wage and enterprise income and in the value of the animal stock was smaller in the MV communities. The differences, however, are not statistically significant.

Table 32. Income trends ($PPP) standard errors in parenthesis

	Wage income	Enterprise income	Animal holdings
Builsa	-27	-5	-19
	(87)	(27)	(138)
MV areas	77	29	62
	(57)	(41)	(137)
Time	65**	84**	183***
	(20)	(28)	(32)
Time*MV areas	-27	-17	17
	(24)	(30)	(60)

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

137. We also tested differential price trends by regressing prices in the logs on time and time interacted. This was done by using a dummy for the MV areas for some prices where complete data are available. Consumer goods are available for sale in few communities and prices were not reported in the community questionnaire. The regression coefficients represent yearly percentage rates of change. The interest in the model is in the interaction between time and the dummy testing in the MV area in regards to whether the price trend is different in the MVs compared to the CV. The only statistically significant difference was found in nominal wages that appear to have grown more slowly

in the MV areas over the last four years, whilst no statistically significant differences were found for the price of fertiliser, goats, and guinea fowl.

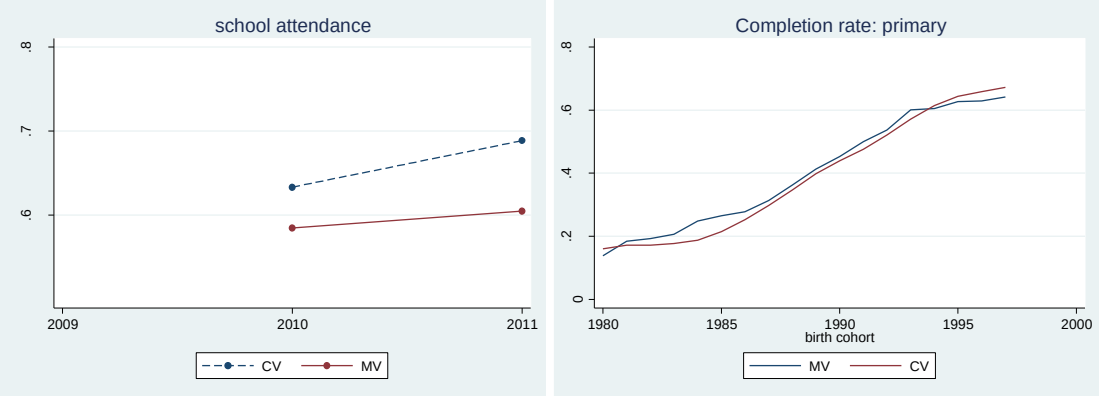
Table 33. Price trends, standard errors in parenthesis

	Male wage	NPK fertiliser	Goats	Guinea fowl
Builsa	0.03	0.01	-0.17***	0.17***
	(0.04)	(0.02)	(0.050)	(0.038)
MV localities	0.55**	-0.39**	0.04	-0.18
	(0.21)	(0.1)	(0.272)	(0.208)

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

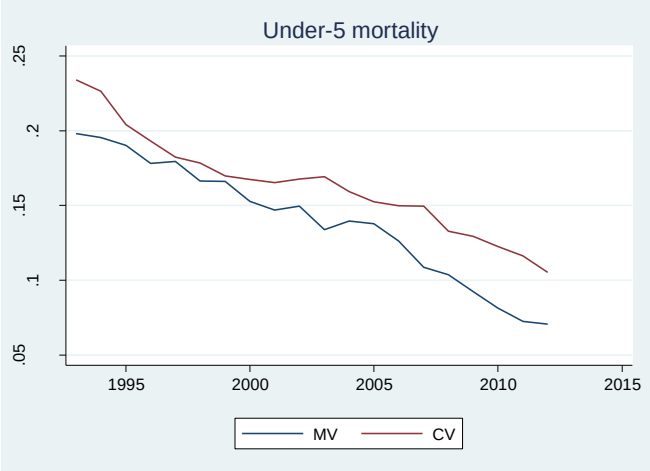
138. The analysis of education trends produces less clear results. In the household survey, respondents reported school attendance in the current and previous year. Whilst there is a sizable and statistically significant difference in attendance rates in primary school at the baseline, the difference is reduced by half and no longer statistically significant in the year before the survey (left chart of Figure 9). We also computed retrospective primary completion rates by birth cohorts (right chart of Figure 9). The birth cohort analysis cannot cover more recent years but the two areas display remarkably similar trends in completion rate up to very recently.

Figure 9. Trends in education indicators



139. In the case of child mortality rates, the differences between MV and CV areas are more interesting and also more difficult to interpret. These rates, built retrospectively using the birth history section of the questionnaire, are displayed in Figure 10 for the 20 years before the survey. There was a remarkable decrease in mortality rates that were cut by half over the period considered in both MV and CV areas. Note also that mortality rates were quite similar in the two areas in the late 1990s, but that they started to diverge in the last decade as the reduction was steeper in MVs compared to CVs. This type of trend is particularly dangerous in DD analysis. If the forces driving the trends operate over the long term and are still in operation, they can easily bias the results. In this particular case a 'programme effect' might be observed even in the absence of the programme.

Figure 10. Trends in under-5 mortality rates in MV and CV areas



140. DD analysis can be affected by the biases described above. It can be observed that case biases could emerge in the analysis of education and particularly health outcomes. The estimation of programme impact can be improved by combining DD with matching, particularly by matching observations using village-level historical trends in the data. In addition, further analysis of the trend data will shed more light on the potential biases and will help establish the proper level of confidence in the results obtained.

5.4 Seasonality

141. The survey questionnaires were administered at different times of the year. For each survey there are gaps of different size between project and control areas: (i) Four-month gap in the community questionnaire; (ii) Eight months for the facility questionnaire; (iii) Two to three months for the household questionnaire; (iv) Three to four months for the adult questionnaire; (v) Two to three months for anthropometric measurements and blood test; (vi) No gap in cognitive and education tests or the expectation tests conducted by the Institute of Statistical, Social and Economic Research (ISSER). Tables 34 and 35 show the percentage of surveys by month in the project and control areas separately.

Table 34. Surveys in the MV areas

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
Community						100 %							
Facilities	30%	60%	10%										
Households					66%	25%	5%	4%					
Adults				42%	40%	16%	1%			1%			
Anthropometry					71%	29%							
Blood tests					68%	32%							
Education tests											90%	10%	

Table 35. Surveys in the control areas

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
Community											100 %		
Facilities										5%	60%	10%	25%
Households								22%	77%	1%			
Adults									92%	7%	1%		
Anthropometry								25%	69%	6%			
Blood tests								32%	52%	16%			
Education tests											90%	10%	

Note: Percentages for the Facility and Community surveys are approximations.

142. In the study area, there is only one rainy season occurring irregularly between June and September, which influences the following variables: (i) Production and consumption patterns; (ii) Morbidity; and (iii) Other behaviours such as, for example, migration. Some of the variables collected by the survey are known to be highly sensitive to seasonal variations such as incidence of diarrhoea and malaria. In addition, even for those variables that are not seasonally sensitive, reporting can be affected by recall bias. For example, respondents' consumption of soft drinks over the previous 12 months may be reported very differently in the dry and the rainy seasons. There is therefore a possibility that some differences observed between project and control areas are the result of the two surveys being administered in different seasons. The questionnaires were designed in such a way to prevent the emergence of a seasonal bias as much as possible. For example, expenditure data were collected with reference to the previous 12 months rather than the previous month as is standard in similar surveys. The agricultural production data were collected with explicit reference to the agricultural year from May 2011 to April 2012. School attendance was reported with reference to the 2011-2012 school year. Despite these efforts a seasonal bias cannot be excluded.

143. Ignoring seasonality in the DD analysis might have disastrous consequences. For example, consider the seasonal patterns of malaria incidence observed by Cairns et al. (2011) in the Navrongo study in the Upper East of Ghana, a region very close to the MV and CV areas. Suppose that, as in the case of the MVP survey, baseline data are collected in May 2002 in project areas (lowest malaria incidence in the year with near to zero incidence) and in September 2002 in control areas (when malaria peaks). Suppose now that data are collected again after one year in May 2003 in both project and control areas. A DD analysis would estimate a negative impact of MV as the control areas are registering a large drop in malaria incidence, though this is simply generated by the seasonal cycle of rains and infection. Note also that repeating the follow-up survey in the same months as in the baseline (i.e. survey in May 2003 in project areas and survey in September 2003 in control areas) would be of little help. The DD estimator would still find a large drop in malaria incidence in control areas. This occurs because there is no regularity in the starting time of rainy seasons and in their intensity from year-to-year. Therefore, all variables influenced by rainfall and seasonal patterns run the risk of being erroneously estimated by a DD analysis. Even variables that are not directly influenced by rainfall and seasons can be erroneously estimated because, as observed above, people's responses to questions can be biased by the time of the year in which the interviews take place.

Figure 11. Seasonal pattern in malaria incidence from Cairns et al. (2011)

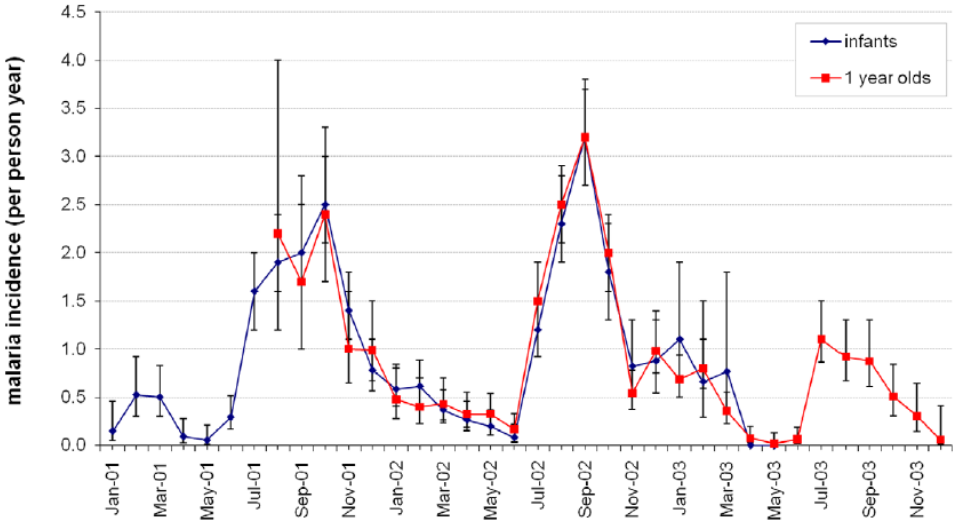


Figure 3. Incidence of clinical malaria in the Navrongo IPTi trial. Incidence of clinical malaria between January 2001 and December 2003 is shown for infants and children 12–23 months of age in the placebo group of the IPTi trial. Error bars indicate 95% confidence interval. Children were enrolled between September 2000– June 2002. Completion of 24 months follow-up ended in June 2004.
doi:10.1371/journal.pone.0018947.g003

144. A number of analyses were conducted using secondary data and a small literature review in order to assess the relevance of seasonality for the outcome variables observed in the study. This work suggested that anaemia, malaria, and self-reported health were likely to be affected whilst anthropometric measurements, income, and consumption were less likely to be affected. The presence of seasonality was tested in the MV and CV data for all the variables for which the presence of seasonality was suspected. This exercise was conducted by running regressions of outcome variables on monthly dummies for the months of May to September and including a control variable for the district surveyed to remove effects related to the survey's geographical roll-out over time. If seasonal effects are found for a variable, it is likely that the difference observed between project and control areas for the same variable is the result of these seasonal effects rather than structural differences between project and control areas.

145. A seasonal pattern in enrolment rates was found: school attendance is lower in the month of July (corresponding to the school break) and increases gradually from September to November. There are at least two explanations for this. The first is censoring. For a given age, children interviewed in November are more likely to have attended school anytime over the previous year. The second is recall bias, which could explain the drop in reported attendance during the school break. The questionnaire asked parents to report school attendance over the school year of 2011-2012, but it is possible that parents interpreted the question differently as, for example, whether the child is currently in school or not.

146. There is a seasonal pattern in time spent on household chores such as collecting wood and childcare. The overall household time spent on these tasks decreases in July and then gradually increases over the following months. This pattern is probably the result of varying degrees of labour use over the dry and rainy seasons and of the varying workload allocation within the family. Among shocks

affecting households, only drought appears to have a seasonal pattern with higher reporting in August and September.

147. There is a strong seasonal pattern in the levels of haemoglobin and related incidence of anaemia. Haemoglobin decreases in the rainy season, reaching its lowest point in July and August after which it increases again. This is consistent with the malaria peak observed in the literature from August to September. There is also an obvious seasonal pattern in using mosquito nets, which is higher in the rainy season compared to the months of May and June. Survey respondents explicitly report not using mosquito nets in the dry season because of the absence of mosquitoes.

148. There is a small seasonal pattern in nutrition indicators. Weight-for-age and weight-for-height are consistently higher in September, which could be related to patterns of food intake after the first harvest. The effect however is small and does not compromise the general comparability of anthropometric data between project and control areas. No seasonal pattern was found in income and expenditure data. These data were collected with reference to the previous 12 months and therefore de-seasonalised. Recall bias seems not to have affected expenditure and production reporting over the previous 12 months. Finally, the cognitive and education tests, survival, and income expectations and preference scores were not affected by seasonal bias because data collection took place simultaneously in all areas in November.

149. The seasonal analysis of the data suggests that many of the characteristic differences between project and control areas are the result of seasonality rather than structural differences between the project and the control communities. The variables affected by seasonality are school attendance, time use, anaemia, and some health-related behaviour such as the use of mosquito nets and incidence of diarrhoea. If these seasonally-affected variables were removed from the balancing test of Section 5.1, there would be only four statistically significant differences, which is well below the acceptable 10% of the sample of variables considered. This in turn suggests that the strategy adopted to select the control communities was valid although the decision to conduct surveys in the project and control communities in different seasons was a very unfortunate one.

150. There are several ways in which the seasonal bias in the data can be assessed and addressed, however none of these procedures can confidently and entirely remove a bias, if present. One option consists of modelling seasonality directly using rainfall data. If the outcome variable response (for example malaria) to rainfall can be precisely estimated using historical data by season, then rainfall data from the project and control sites would allow an estimation of seasonal bias. Unfortunately, similar data are unlikely to be available. A second option consists of estimating seasonal bias by employing a Oaxaca-Blinder decomposition. If the outcome variable of interest can be modelled and no determinants are omitted, then the seasonal bias can be measured directly. This type of analysis is based on the assumption that no determinants of the outcome variable (observable or unobservable) are omitted from the model, which is a strong assumption. A third option consists of conducting the estimation separately for subgroups of the population, some of which are known to be unaffected by seasonality. For example, malaria appears to be seasonal for infants but not for older children. If a similar pattern is found, this provides evidence of a seasonal bias. A fourth option consists of exploiting the survey timing and comparing groups of households interviewed at different times of the year. As the time gap between the groups grows, the difference in the outcome variable should also increase following the expected seasonal trend, which thus provides seasonal bias evidence. Finally, an overall assessment of the size of the bias can be obtained by comparing the DD impact results calculated separately by using the unadjusted data, data adjusted by propensity score

matching, data adjusted by a Oaxaca correction, and data limited to subgroups of the whole population (see Appendix C for further information about this procedure).

151. As the project was not randomly allocated to a locality and because there is no time overlap between the surveys conducted in project and control villages, it is extremely difficult to define with certainty to what extent the differences observed are the result of seasonal bias or of underlying structural differences in the populations. Ultimately, the evaluation team will have to rely on well-reasoned assumptions in their efforts to correct for the seasonality issue, alongside employing all of the strategies outlined above in order to detect the size of the bias. Additional information for testing the presence of a bias will become available with the first and second follow-up rounds. For example, variables that are known not be affected, or poorly affected, by the project should show no differences at follow-ups because they are simultaneously conducted in the MV and CV areas. For this reason and for the general uncertainty regarding the size of the seasonal bias it is recommend that the survey continue as planned in the next follow-up rounds even for those variables, such as use of mosquito nets and malaria incidence, whose validity is more questionable.

5.5 Data quality and Benford's Law

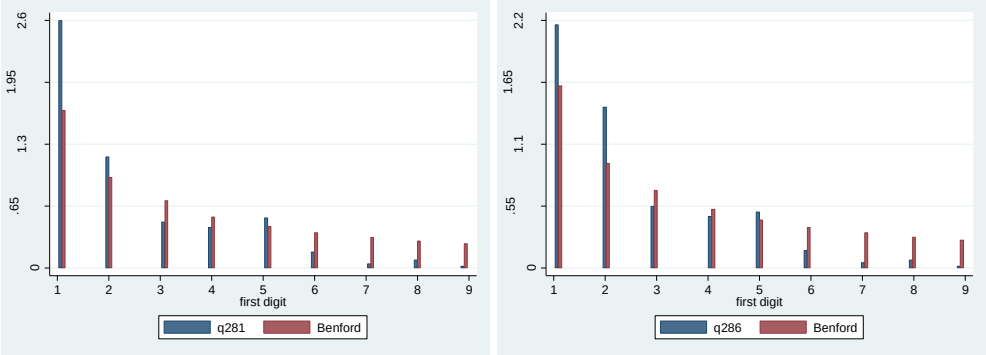
152. The survey data suffer problems typical of data collection in developing countries: age heaping, unreasonably large values of goods purchased and produced, and standard deviations larger than the typical norm (e.g. nutrition). There is no obvious indication that these problems are larger in the MVP survey than in comparable surveys conducted in Ghana. To better assess the quality of the data we decided to check their reliability using Benford's Law. First digits of many social as well as natural phenomena do not follow a uniform distribution. A larger proportion of numbers have one as first digit and then two, three to nine in decreasing proportions (Benford's Law). This regularity has led some researchers to use the distance between observed data and the theoretical Benford's distribution as a tool for detecting fraudulent data fabrications. Judge et al. (2009) apply Benford's Law to a number of datasets from developing countries and what follows is an application of the metrics and statistical tests they developed.

153. We applied Benford's Law to purchases and own consumption figures from the expenditure file and to quantities of agricultural output produced from the farm income data. These quantities were looked at in a comparative way by calculating the same metrics and statistical tests for comparable datasets collected in Ghana: the GLSS collected by the GSS in 2005/2006 and the Yale/ISSER panel dataset of 2009. This type of analysis can only be applied to variables whose numbers are reported in large amounts and whose first digits do not follow any particular distribution. For example, first digits of age data are largely determined by survival patterns. Additional variables that could be tested using these datasets include the number of animals owned and possibly land size.

154. Differences between the observed distribution of first digits and the theoretical Benford distribution were looked at. These differences are tested using a chi-square test and a Kuiper's test. Unfortunately, the formulae for the tests calculations included a term for the size of the sample. As a result, the size of the test simply increases with the sample size. Hence, whilst the test is useful to assess the conformity to the Benford distribution it does not measure the distance from the distribution. It would be an error to interpret a larger value of the test as a larger distance from the theoretical Benford distribution. For this reason, the size of the test cannot be used to make comparisons across surveys to say, for example, that survey A is more accurate because Kuiper's test is 1,000 whilst survey B is 1,500. In order to make comparisons across surveys, we also report a

number of distance measurements between the observed distribution and the Benford distribution. These are the M and D distances in Tables 36 and 37 that will be used to evaluate the surveys in a comparative way.

Figure 12. Observed and Benford distributions compared: food purchases and own consumption



Note: Calculated using methodology in Judge et al. (2009) stata do-files available with the authors.
Note: q281 and q286 are the names employed in the survey data for the value amounts of food purchases (q281) and the value amount of consumption of foods produced by the household (q286).

155. Table 36 compares the tests of equality of the two distributions of the MV data with data collected by the ISSER survey in 2009 and by GSS in 2005. Only observations from rural households from the Northern, Upper East, and Upper West regions were considered in order to make them more comparable with the MV datasets. Similar to the MV datasets, the expenditure modules of ISSER and GSS also employ a variety of local units of measurement for the item purchased. The recall periods used however are very different. The GSS employed diaries with repeat visits and a recall of 15 days, whilst the ISSER questionnaire employed a 30-day recall. The MV questionnaire employed a recall consisting of the quantity purchased and consumed in a typical month for the months it was purchased or consumed.

Table 36. Quality analysis of expenditure data based on Benford's Law of three different datasets

	Obs	M distance	D* distance	Chi-square	Kuiper's test
Purchases					
GSS 2005	-			-	-
ISSER 2009	12,585	0.093	0.124	2,110.2***	16.9***
MVP 2012	29,298	0.201	0.171	7,414.0***	36.1***
MV areas	9,282	0.170	0.183	2,272.8***	19.3***
CV areas	6,108	0.172	0.187	5,181.2***	30.5***
Own consumption					
GSS 2005	4,769	0.104	0.146	1,095.0***	11.3***
ISSER 2009	3,679	0.044	0.092	455.7***	7.6***
MVP 2012	19,107	0.101	0.151	4,222.0***	26.9***
MV areas	6,108	0.117	0.156	1,516.8***	14.9***
CV areas	12,999	0.115	0.153	2,794.2***	22.3***

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%
Note: Calculated using methodology in Judge et al. (2009) stata do-files available with the authors.

156. In all cases, the Benford distribution and the observed distributions are very different at 1% statistical significance. For comparative purposes, the maximum distance (M) and the Euclidean distance (D*)

were used. Based on these two measurements, the MV data appear to be less accurate than the ISSER data but of comparable quality to the GSS data. More importantly, the different measure is calculated separately for the MV and CV sites of the MVP survey because the surveys were conducted at different times of the year and by slightly different teams of enumerators. No differences in the quality of MV and CV data are found based on these measures. The analysis of agricultural production data leads to a similar conclusion. By an application of Benford's Law, the data collected by the MVP do not appear to be of inferior quality to those collected by ISSER or GSS nor do large differences appear between data collected in the project and control communities.

Figure 13. Observed and Benford distributions compared: agricultural production

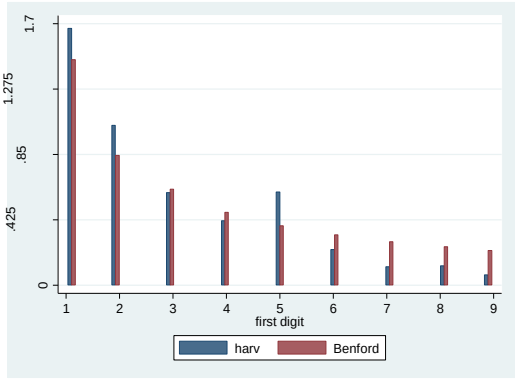


Table 37. Quality analysis of expenditure data based on Benford's Law of three different datasets

	Obs	M distance	D* distance	Chi-square	Kuiper's test
Harvested q					
GSS 2005	8,254	0.028	0.041	257.2***	4.2***
ISSER 2009	3,327	0.033	0.070	223.5***	5.6***
MVP 2012	7,528	0.045	0.086	761.6***	9.7***
MV areas	2,651	0.048	0.084	280.7***	5.9***
CV areas	4,877	0.057	0.092	509.9***	7.7***

*Difference statistically significant at 10%, ** significant at 5%, *** significant at 1%

5.6 Post-hoc power calculations

157. The sample size of the baseline household survey was set based on power calculations looking at three main outcome indicators: per-capita expenditure, height-for-age Z-scores of children under 5, and maths test scores among children of primary school age. These calculations led to the selection of 35 clusters, with 20 households per cluster, for the project group and for each of the comparator groups. The power calculations conducted at the design stage were based on parameters obtained from available household surveys of Ghanaian households. The data collected at the baseline offer the opportunity to validate the parameters used at the design stage and to recalculate the power of the sample with respect to the same indicators.

158. The analysis of the baseline data reveals that the parameters employed at the design stage were largely similar to those obtainable from the baseline data (see Table 38). At the baseline, the sample size turned out to be slightly larger because it was purposely decided to oversample households in order to account for the occurrence of attrition. This had the effect of increasing the power of the sample with respect to estimates at the design stage. The standard deviations of the variables at the baseline are very similar to those assumed at the design stage and slightly larger only in the case of

per capita expenditure. This also had the effect of increasing the power of the sample with respect to the estimate at the design stage.[40] Finally, the intra-cluster correlation coefficients are much larger at the baseline than what were assumed at the design stage, though per capita expenditure is an exception.[41] Larger intra-cluster correlation coefficients have the effect of reducing power.

Table 38. Parameters employed in the post-hoc power calculations

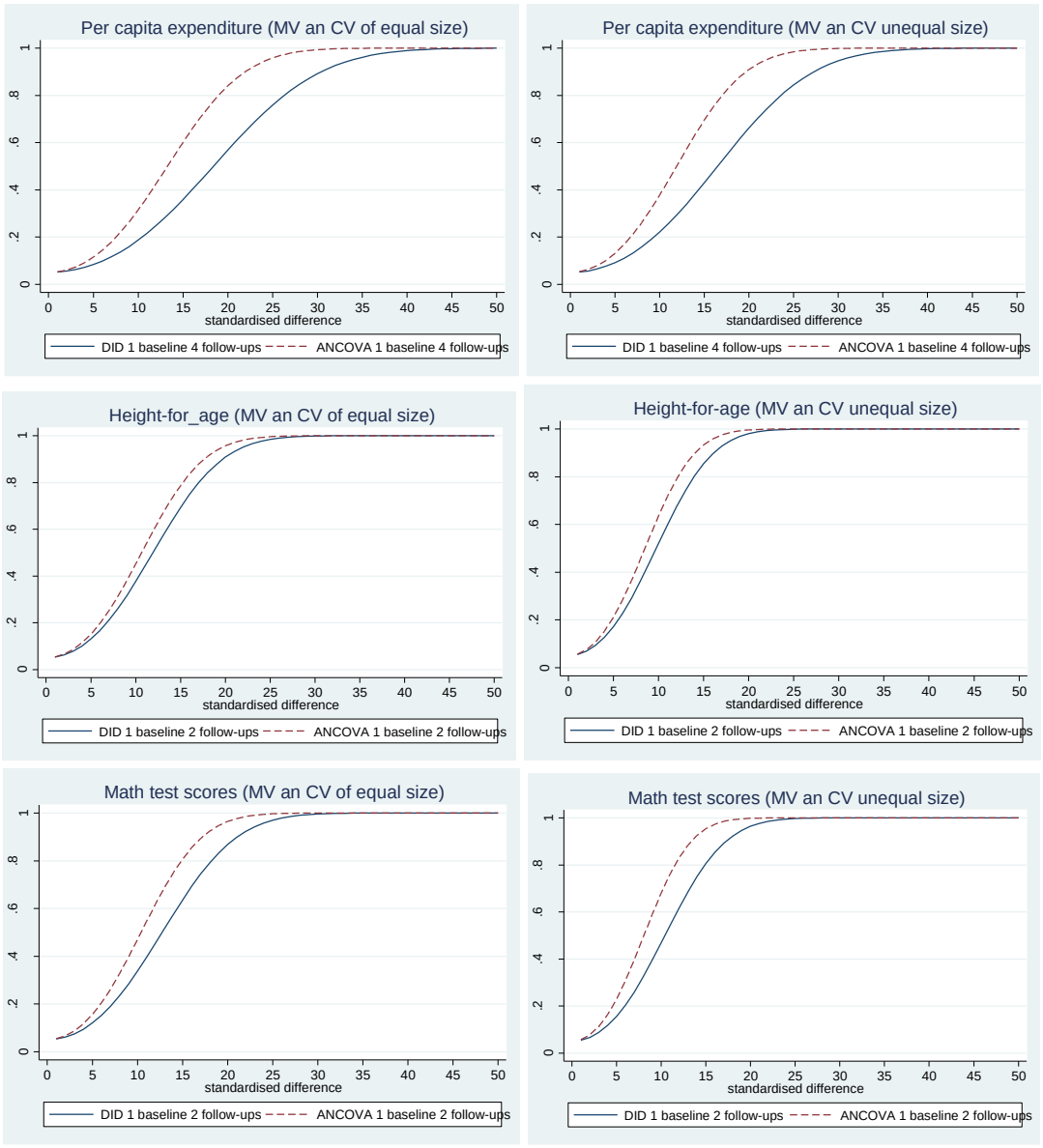
Variable	Cluster size	Ratio mean to standard deviation	Intra-cluster correlation coefficient	Autocorrelation coefficient
Per capita expenditure	21	1.5	0.17	0.40
Height-for-age Z-score	19	0.7	0.10	0.70
Easy maths test score	51	2.4	0.08	0.50

159. It is difficult to predict the overall change in the power of the sample by simply looking at Table 38, as while some parameters suggest an increase in power others suggest a decrease. Hence, it was decided to perform the same power calculations conducted at the design stage over the same indicators and then compare the results. Two study designs are considered: a DD design and an analysis of covariance (ANCOVA) design consisting of one baseline and two follow-ups with the exception of per capita expenditure for which four follow-ups are available in both cases. The autocorrelation parameters for the power calculations are reported in Table 38 and are the same employed at the design stage. The results of these power calculations are shown as power against standardised differences (Figure 14). The charts on the left side calculate power when the MV sample is compared to a CV sample of the same size whilst the charts on the right side calculate power when the MV sample is compared to the combined sample of near and far CV sites.

[40] For comparative purposes, the ratios of means to standard deviations for the three indicators in Table 38 at the design stage were set at 1.7 for per capita expenditure, 0.7 for height-for-age Z-scores and 2.0 for math test scores.

[41] For comparative purposes, the intra-cluster correlation coefficients for the three indicators in Table 38 at the design stage were set at 0.23 for per capita expenditure, 0.03 for height-for-age Z-scores and 0.02 for math test scores.

Figure 14. Power against standardised effect size



160. The results show that in all cases the detectable difference at 80% power is below the critical threshold of 0.2 standard deviations. The sample will be able to detect 'small' impacts. In the case of per capita expenditure the study will exploit four follow-ups and will be able to detect an impact of 0.15 standard deviations, which is roughly equivalent to an increase in per capita expenditure by 11%. In the case of height-for-age Z-scores and of maths test the study will be able to detect an impact in the range of 0.1-0.15 standard deviations. The power of the samples turned out to be larger at the baseline in the case of per capita expenditure (detectable difference was 0.2-0.3 standard deviations at the design stage) but slightly lower in the case of height-for-age and easy maths test scores (detectable differences of about 0.1 and 0.08-0.10, respectively).

161. In conclusion, the power calculations performed using baseline parameter data do not differ substantially from those conducted at the design stage. The study is found to be better powered to assess the impact on poverty but less powered to observe impact on nutritional and educational

outcome than was originally thought. However, it should be acknowledged that the study has in all cases sufficient power to detect impacts below the critical threshold of 0.2 standard deviations, which is often classified as 'small' impact. The study is therefore sufficiently powered to detect a small impact on the three indicators considered.

6. Summary and conclusions

162. The baseline report is a summary of the data collected during 2012 and early 2013. It provides an overview of the MVP area at the start of the project and summarises key variables across poverty, health, education, etc. that will be used to measure impact at the end of the evaluation in 2016/17. For this reason, this report is mainly summative and focuses on ensuring that the data are consistent and of high quality whilst observing any interesting or unusual deviations (e.g. between MV and CV datasets, through comparisons with national datasets, and other known phenomenon such as the Engel Curve).

6.1 Characteristics of the study population

163. The overarching goal of the MVP is to make progress against the MDGs in Northern Ghana. Unsurprisingly, income poverty and inequality is much higher in the study area than compared with the rest of the country, and indicators for educational attainment and maternal health are all lower. The overall picture is one of a deprived area where economic, education, and health conditions are very poor. Nonetheless there are some surprising facts. The under-five mortality rates and undernutrition rates are comparable to national averages, and gender differences in education indicate under-enrolment and participation by boys.

164. In terms of household characteristics, the average size in the study population is seven members. Women head around 10% of households and around 20% are polygamous. About 1 household in 10 hosts a migrant, whilst around half of households have a member who has temporarily migrated away. Poverty rates in the MV sites are very high (around 60%) and, surprisingly, inequality indicators are higher than the rest of the country. Nonetheless, despite a large proportion of imputed expenditure devoted to food, households are not able to meet their food requirements as some 80% of households report months when they did not have enough food to meet their family's needs. There is also a strong seasonal pattern to food availability with only one cropping season.

165. The majority (90%) of households across the study area are farmers, cultivating an average of three hectares of land scattered across an average of three different plots. Most farming is subsistence based, and only around 20% of agricultural produce is sold on the market. Farming in the area operates under risky circumstances, with all households having been affected by a 'shock' of some kind in the preceding year. Very few households save and the most common form of 'savings' consists of animal holdings.

166. In terms of education, primary school attendance is lower than in comparison to the rest of Ghana, but not by a large margin. Attendance at JHS and SHS is very low and according to the surveys there are more girls than boys in school at all grades. The quality of schooling is poor, as is apparent in the results of the maths and English tests conducted by the survey team.

167. Mortality rates in the study area are high by international standards, but not in comparison with the rest of Ghana and are considerably lower than those calculated by the DHS in the northern regions in 2008. This may be because mortality rates have been declining since the figures were reported in 2008 or because the DHS estimates are valid at the national level (i.e. regional level estimates tend to be more inaccurate). Undernutrition rates are high and very similar to malnutrition rates observed for the whole of Ghana by the DHS in 2008.

168. There are extensive social networks across communities in the study area with around 80% of households reporting having relatives or friends living in another village and approximately 50% of those relatives or friends having asked for or provided help in the previous 12 months.

6.2 Characteristics of the data – balance tests

169. The DD evaluation design rests on the assumption of parallel trends. For comparisons in the changes in the outcomes in the project and control group to be valid, the pre-interventions trends of the outcome variables should be similar. For many outcome variables however differences in the levels may suggest differences in the trends as well. For example, the demographic transition to smaller families is not a linear process and differences in the average levels would suggest differences in the trends as well. The selection of control communities was done by matching rather than by random selection, which was deemed infeasible for the MVP. To assess whether there are statistically significant differences between project and control villages, there is a list of variables against which comparisons are made between both the size of the difference and the statistical significance.

170. For the household data, of the 52 variables considered, 11 were found to be different at a 10% statistical level of confidence, representing around 20% of all variables (which is similar to the differences found in the community-level data). A large number of differences are found among seasonally sensitive variables (e.g. episodes of diarrhoea), and some differences could be a reflection of differences in socio-economic characteristics of the two areas. The faraway control communities are not more different than nearby communities, as the number of statistically significant differences is the same for both groups.

171. DD analysis also rests on the hypothesis of parallel trends. The trends show that the increase in wage and enterprise income was smaller in MV communities, whilst the reduction in the value of animal stock was larger. The differences were not however statistically significant. For price trends, nominal wages appear to have grown more slowly in MV areas over the last four years with no statistically significant differences in the price of fertiliser, goats, and guinea fowl.

6.3 Characteristics of the data – seasonality

172. The survey was administered at different times of the year with gaps of different sizes for the same survey between project and control groups. There is one rainy season occurring irregularly between June and September, which influences production and consumption patterns, morbidity, and other behaviours such as migration. Some of the variables collected by the survey are known to be sensitive to seasonal variations, such as the incidence of diarrhoea and malaria. For other variables, reporting can be affected by recall bias.

173. The presence of seasonality in the data was tested by running regressions of dependent variables on month dummies from May to September with a control variable for the district surveyed in order to remove effects related to the geographic roll-out of the survey over time. The analysis suggests that many of the differences in characteristics between project and control areas observed are the result of seasonality rather than structural differences between project and control communities. The variables affected are school attendance, time use, anaemia, and some health related behaviour such as the use of mosquito nets and incidence of diarrhoea. This suggests that the strategy adopted to select the control communities was valid.

6.4 Characteristics of the data – quality

174. The data suffer from typical data collection problems in developing countries, such as age heaping, unreasonably large values of good purchased and produced, and standard deviations larger than the typical norm. There is however no obvious indication that these problems are larger in the MVP survey than in comparable surveys conducted in Ghana. The evaluation also tested the reliability of the dataset by using Benford's Law. In all cases, the Benford distribution and the observed distributions are very different at 1% statistical significance. Based on these tests, the MV dataset appears to be less accurate than the ISSER data but of comparable quality to the GSS data. Also, no differences have been found in the quality of MV and CV data based on these measures.

175. Overall, the strategy adopted to select the control communities is valid, and the quality of the dataset appears to be no less accurate than comparable datasets collected in Ghana. The time difference in the collection of baseline data in the project and control communities does however mean that some variables are particularly sensitive to seasonal differences.

7. References

Bhalotra, S., and C. Attfield. 1998. Intrahousehold resource allocation in rural Pakistan: A semiparametric analysis. *Journal of Applied Econometrics* 13 (5):463-480.

Cohen, J. 1988. *Statistical Power Analyisis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Duflo, E., M. Kremer, and J. Robinson. 2011. Nudging Farmers to use Fertiliser: Theory and Experimental Evidence from Kenya. *American Economic Review* 101 (6):2350-2390.

Fisher, I. 1930. *The Theory of Interest, as Determined by the Impatience to Spend Income and Opportunity to Invest It*. New York: Macmillan.

Guyer, J. I. 2004. *Marginal Gains: Monetary transactions in Atlantic Africa*. Chicago: The University of Chicago Press.

Kedir, A., and S. Girma. 2007. Quadratic engel curves with measurement error: Evidence from a budget survey. *Oxford Bulletin of Economics and Statistics* 69 (1):123-138.

Lewbel, A. 2008. Engel Curves. In *The New Palgrave Dictionary of Economics*, edited by S. N. Durlauf and L. E. Blume. Houndmills Basingstoke: Palgrave Macmillan.

Udry, C., and H. Woo. 2007. Households and the Social Organisation of Consumption in Southern Ghana. *African Studies Review* 50 (2):139-153.

White, H. 2009. *Theory-Based Impact Evaluation: Principles and Practice*, 3ie Working Paper 3, New Dehli: International Initiative for Impact Evaluation (3ie).

World Bank. 2011. *Republic of Ghana Tackling Poverty in Northern Ghana*, *PREM 4/AFTAR*. Washington DC: The World Bank.

APPENDIX A. TERMS OF REFERENCE

PO 5603 MV-EVALUATION: TERMS OF REFERENCE

Title:	Impact Evaluation of a New Millennium Village in Northern Ghana

1. Introduction

1.1. The UK government's Department for International Development (DFID) manages Britain's development assistance to poor countries and works to eradicate extreme poverty. We are led by a cabinet minister, one of the senior ministers in the government. This in itself is a sign of how determined the UK government is to tackle poverty around the world. Guided by these principles, DFID works across the world on a bilateral basis with partner countries, with multilateral organisations, and with civil society.

1.2. The Government of Ghana and DFID will be working with the Millennium Promise Alliance (MPA) to implement a Millennium Village (MV) in Northern Ghana. The MV project will commence in late 2011 and will last five years. The MV model is already being implemented in a range of sites across Sub-Saharan Africa, where it is now entering a second five-year phase.

1.3. The MV model provides an integrated package of interventions to lift a rural community out of poverty. Its central hypothesis is that a local 'big push' addressing the most immediate capital deficiencies in communities and households is a necessary condition for reaching a threshold required to move towards local resilience and self-sustaining economic growth. Key to this is improved agricultural productivity and market development, enabling people in rural areas to save and accumulate wealth, stimulating investment and diversification into non-farm work.

1.4. The MV projects across Africa have set up comprehensive Monitoring and Evaluation (M&E) systems. These are used to continually assess progress and adapt implementation mechanisms. The datasets produced have also fed into MV reports, including on results achieved. However, there is a noticeable gap in evidence of the model's overall effectiveness. A key MV report of results achieved was based on before-and-after analysis within the MV sites[i], leading to criticism of the results attributed to the MVs and the lack of independent rigorous evaluation[ii]. DFID has agreed with MPA that funding for a new MV in Northern Ghana will be accompanied by such an independent evaluation, to provide robust evidence on the effectiveness of the MV approach.

2. Objectives

2.1. DFID wishes to invite suitably qualified organisations to implement a robust independent evaluation of the new MV in Northern Ghana. The evaluation will cover

up to a 10-year period – subject to programme renewal – to answer evaluation questions of importance to the Government of Ghana, its Savannah Accelerated Development Authority (SADA), local stakeholders, DFID, and the international development community.

2.2. The evaluation will cover the costs of generating, analysing and quality assuring data, producing reports, and widely disseminating the results of the evaluation. The independent evaluation will build on, expand and validate the MV project's own M&E of the MV site and their selected comparison site. It will include establishing baselines, ongoing evaluation during the implementation phase and, subject to further agreement, continued evaluation after completion of the five years of direct implementation by the MV project.

3. Recipient

3.1. The recipient is DFID, with the project being managed by the DFID Ghana office.

4. Scope of Services

4.1. Appropriate methodologies will be used to answer the four key questions underpinning the evaluation of the MV in Northern Ghana:

i. Does the MV *deliver on promises to reach the Millennium Development Goals* (MDGs) within the MV site?
ii. Are the positive impacts of the MV *sustainable* after direct implementation of the MV project has ended?
iii. Is the MV intervention package cost-effective in the results it achieves, compared with possible alternatives?
iv. What *externalities or spillover effects* does the MV generate, and do they significantly add to or detract from the positive impacts that might be achieved within the MV site?

4.2. Besides the four main evaluation questions listed above, key stakeholders have also raised other issues that need to be explored in the evaluation. In particular, the methodology and evaluation will aim to also examine:

a. Does the MV package *empower disadvantaged or marginalised* groups (e.g. females, the disabled, or the elderly)?
b. Does the MV achieve *additional benefits arising from synergies* across implementation of an integrated package of interventions?
c. Does the MV address common issues relating to *agriculture, infrastructure, or social and economic concerns*?[42]

[42] Examples of questions relating to agriculture, infrastructure, and social and economic concerns are in the field visit report included in the list of documents.

4.3. Given the aim of evaluating the MV model as it will be applied in Northern Ghana, the evaluation methodologies employed will not require a change in MV implementation. This is likely to preclude the use of randomised control trials, due to the nature of the MV's integrated package of interventions across a single site. However, proposal of any evaluation methodologies, including randomised approaches, will be considered if they are feasible, cost-effective, and able to answer the key evaluation questions.

4.4. At this point in time, the Evaluation Advisory Group for this independent MV evaluation considers that a difference-in-differences approach with mixed methods is the most likely approach to be able to meet the criteria.[43]

4.5. The independent evaluation will work with the MV project to identify appropriate comparison sites, and may need to suggest additional variables to be used in the village matching process.[44]

4.6. Surveys at the MV site and comparison sites will take place in year 1 (baseline) and at least twice more during the five-year MV direct implementation period.

4.7. Spillover effects in areas adjacent to the MV (and possibly beyond) will need to be assessed. Various methods could be employed for this, but must be cost-effective given the primary emphasis on evaluation of impacts within the MV site.

4.8. DFID's funding to the MV in Northern Ghana will be subject to a mid-term review in year three. This will determine if there is sufficient evidence of progress against its objectives to justify completing the full funding to year five. The independent evaluation will play an integral role in this mid-term review, providing a report on progress, assessing cost-effectiveness, and producing a cost-benefit analysis of the MV in Northern Ghana based on the evidence available at that point.

5. The Requirements

5.1. The evaluation must be carried out by researchers with a recognised international reputation and practical experience of rigorous impact evaluation. The evaluation must reflect the local context. It must be independent, robust, and credible. Findings of the evaluation should be published in standalone reports and through peer-reviewed journals.

5.2. The independent evaluation is being funded, sourced, and delivered separately from:

- The main project under which DFID will fund the implementation of a new MV in Northern Ghana.

[43] A DFID team visited the proposed MV site in Northern Ghana and compiled a description of the site and a potential evaluation approach. Selected sections of the Visit Report are included in the accompanying documents.
[44] The field report contains the current village matching checklist used by the MVP.

- The MV project's own internal arrangements for monitoring and evaluation. These will continue during the MV project period. They are essential for the MV's own management and implementation, and will also provide a major portion of the data required for this independent evaluation.

5.3. Where data generated internally by the MV project are used, independent verification is required, if necessary on a sampling basis. The independent evaluation will need to verify the accuracy of surveys conducted by the MV project. The independent evaluation will be responsible for choosing the scale of surveys and the degree of sampling required, but methodologies must comply with generally accepted best practice. The independent evaluation will also review all survey instruments before they are sent to the field.

5.4. Additional survey modules or data collection methods may be required to address the key evaluation questions listed above.[45] The independent evaluation may need to work with the MV project to include additional modules in their surveys.

5.5. All findings, datasets, and methods for the evaluation component project must be published and made available to allow researchers to replicate findings. Publication in peer-reviewed journals should be an objective.

5.6. Participation will be expected in various fora, including international and national conferences, particularly in latter years as evidence emerges. This will require high calibre expertise in presenting and debating findings. Costs of participation in such events will be borne by DFID or other parties.

6. Constraints and Dependencies

6.1. DFID Ghana will provide a grant totalling $18.1 million USD over five years for implementation of the new Millennium Village in northern Ghana. This includes resources for the implementation of the MV, as well as technical support required to run the MV project's own M&E systems. The scale of the independent evaluation of the MV project will need to reflect the size of the MV, the degree to which the MV project's own M&E systems can be used and the extent to which their data will need to be validated, the need for any comparison sites in addition to the single comparison site to be selected and monitored by the MV project, and the intended 10-year period of the evaluation. The timeframe for the initial provision of independent evaluation services will be for five years, but the evaluation framework that is designed should be for a full 10-year period.

6.2. The evaluation must remain independent of the MV project's own M&E processes but, at the same time, the evaluation team must work closely with – and can expect full cooperation from – the MV project, including the team working specifically on

[45] A draft of the current survey tools that are used by the MVP is included in the accompanying documents.

M&E for the northern Ghana site, and associated MV organisations.[46] Consistency is crucial between information collected from within the MV and MV-comparison site led by the MV project, and any additional comparison sites that might be led by the independent evaluation. It may be possible to contract the MV project's M&E resources to carry out data collection in additional comparison sites. It may or may not be possible to utilise M&E resources associated with the MVs during the five-year period after direct implementation of the MV in northern Ghana ends.

6.3. There are numerous factors that could have implications for the independent evaluation. For instance, the migration of households into and out of the MV site, and exogenous shocks within the MV site, nearby, or at a national scale. Such problems need to be considered and mitigating actions proposed; for instance, maintaining a statistically valid sample size in the MV and comparison sites will be crucial.

6.4. We do not want to be overly prescriptive on staffing arrangements but expect the evaluation team to put forward a highly experienced small core team of international and national experts, and a network of local field workers, who will be present at the site during key stages. It is also natural to expect a turnover of personnel during the life of the evaluation. Plans and mitigation measures need to be outlined.

6.5. The site is situated in a remote part of northern Ghana, two hours drive from Tamale, the capital of Northern Region (which itself is 10 hours by road, or a 75 minute flight, from Accra). Local access is via basic non-paved roads. Movement across the middle of the proposed site can become restricted in the rainy season when the White Volta River floods, which is why a portion of the site is referred to locally as "the overseas."

6.6. In addition to developing a strong working relationship with the MV project at the site and with MV organisations outside Ghana, the evaluation team will need to engage with other stakeholders. For instance with local communities, district and regional officials, the SADA, national government agencies such as the National Development Planning Commission and the Ghana Statistical Service, and other organisations providing and assessing the impact of external assistance to the area (for example, the Millennium Challenge Corporation, CARE, and IPA/JPAL).

7. Reporting

7.1. The independent evaluation will report regularly to DFID Ghana's MV Adviser.

7.2. Annual reviews of the independent evaluation will be conducted by DFID, which will require full cooperation from the independent evaluation team, including providing an annual progress report against the logframe. These annual reviews will be determined by DFID's internal reporting requirements and may not fit with the schedule of MV surveys.

[46] Such as the Millennium Promise Alliance based in New York, the Earth Institute at Columbia University, and the MDG Centre for West and Central Africa based in Mali.

7.3. An Evaluation Advisory Group, organised by DFID, will guide the strategic direction of the independent evaluation, signing-off on key reports and outputs. This Advisory Group will include representatives of DFID, Government of Ghana, the MV project team, and other key stakeholders. The Advisory Group will play a key role in agreeing the final design for the independent evaluation, and is expected to meet at least before and after each major survey event (including initial establishment of baselines).

8. Timeframe

8.1. The independent evaluation will be designed for a 10-year period, to allow for assessment of sustainability of the MV's impacts, but will be contracted initially for a five-year period.

8.2. The five-year MV intervention is scheduled to start its set-up phase in late 2011, with a detailed design phase of up to six months. All subsequent interventions will be sequenced according to the needs of local circumstances, as determined by the MV project. The MV project's own M&E, establishing detailed baselines, will commence during the design phase.

8.3. The parallel implementation of the Millennium Village and the independent evaluation is critical. Therefore, thorough baselines need to be established very rapidly. Major MV interventions are likely to start in the first few months of 2012. The independent evaluation needs to finalise its approach, identify survey locations and methods, and commence validation of MV baselines and/or establish additional data collection early in 2012. An indicative initial timeline is outlined below.

Date	Output
w/c 3 October	Pre-bid workshop. A half-day conference will be held in East Kilbride, Scotland, organised by DFID, for the MV project to outline their approach to implementation and to M&E, and to enable potential bidders to ask questions of the MV team and of DFID, for instance on survey methods and how new modules could be incorporated.
7 November	Deadline for bid submission.
w/c 21 November	Notification to all candidates of ITT outcome. (NB there is a chance bidders are expected to be invited for follow-up interviews in the up to this point)
December 2011	Contract signed. Initial design of the evaluation commenced.
By end January 2012	Detailed design agreed with the Evaluation Advisory Group.
February 2012	Baseline field activities completed.

8.4. In view of the long time horizon and to allow for changes during the lifetime of the contract, annual review points will be planned. The initial evaluation contract will be let for a period of up to five years in the first instance, and will include break points at

the end of Year 1 and Year 3. Progression from one period to the next will be subject to the satisfactory performance of the Service Provider (SP), the continuing requirement for the services, and agreement on work plans and budgets for the following period.

8.5. At the end of Year 5 DFID will review the requirement, the performance up to that point, and the future scope – to determine whether the independent evaluation should continue to be conducted by the service provider. The contract could then be extended for a period of up to five years, with timing of break points for that extension agreed at that time.

9. Outputs

9.1. The independent evaluation will produce the following outputs:

I. An initial design document within the first six weeks of contract exchange, outlining features of the proposed evaluation framework including key evaluation questions, methodologies to be employed, selection of comparison sites, and ways of working with the MV project and other key stakeholders. Key critics of the MV approach will be consulted on proposed evaluation design options before they are finalised.

II. Baseline surveys completed within the first six months of the implementation of the MV project.

III. Annual Progress Reports, based on DFID's logframe for the independent evaluation, to fit into DFID's internal reporting schedule.

IV. After each survey round, an initial report on evaluation approaches and data issues, and a detailed report following analysis of the data and other information.

V. Mid-term report on the northern Ghana MV, assessing cost-effectiveness, and a cost-benefit analysis based on the evidence available at that point.

VI. 'Final Report' on the northern Ghana MV in Year 5, including answers to the key evaluation questions. A separate, easily understood summary of the evaluation findings.

VII. Data and reports available in the public domain, as quickly as possible.

10. DFID Coordination

10.1. The DFID Ghana MV Adviser will be the direct point of contact in DFID for the independent evaluation, and will arrange meetings of the Evaluation Advisory Group.

11. Background

11.1. Ghana has succeeded in reducing the national rate of poverty from 52% in 1992 to less than 29% in 2006.[iii] This national-level improvement, however, has not been spread evenly. The dry northern savannah in particular experiences persistently high levels of poverty, estimated to be 69% in 2006.[iv] There have been concerted efforts for decades to reduce the stubbornly high rates of poverty in the North[v] but with little success.[vi] The region exhibits the characteristics of what Jeffrey Sachs calls a 'poverty trap' deriving from a paucity of various forms of capital.[vii] The Government of Ghana acknowledges the particular challenges faced by the North, and in 2010 created the semi-autonomous Savannah Accelerated Development Authority (SADA).[viii] The associated SADA Strategy, 'A Sustainable Development Initiative for the Northern Savannah', emphasises "transforming the northern Ghanaian economy and society into a regional nexus of increased productivity of food and a buffer against persistent droughts and sporadic floods."[ix]

11.2. Sachs's ideas for tackling the 'poverty trap' have been taken forward in the form of Millennium Villages (MV's), through the non-profit organisation Millennium Promise.[x] There are currently 12 MV sites being implemented across Africa, assisting communities to lift themselves out of extreme poverty. This is a 'big push' approach, providing an integrated and intensive programme of support and community development to people within a defined area, seeking to show how the Millennium Development Goals (MDGs) can be achieved by 2015, even in very poor rural areas of Africa.

11.3. The first MVs commenced in 2006.[xi] Their average results are reported as including a seven-fold increase in the use of bed nets among children, maize yields having tripled, and access to improved drinking water higher by 50 percentage points.[xii] However, the MVs have been subject to criticism, particularly related to the lack of rigorous independent evaluation of their impact. For instance, some results reported for MVs based on before-and-after comparison were found to have occurred to a similar degree in other sites within the same country.[xiii] Critics suggest that it is unsurprising that channelling significant resources to a relatively small population will have some beneficial impact. Key questions, however, are around the **cost-effectiveness** and the **sustainability** of this approach. For instance, could the impacts achieved at MV sites be achieved at a lower cost through alternative approaches? And are the impacts sustained once the substantial pulse of increased resources to the area comes to an end? This independent evaluation project aims to provide evidence to help answer these questions.

11.4. The MV would represent an innovative approach to addressing the chronic poverty that afflicts North Ghana. It fits well with DFID's increased emphasis on innovation and on achieving real development results.[xiv] And it would be in line with DFID Ghana's new Operational Plan that proposes increased focus on the poor north of the country.[xv] The proposed MV is being closely coordinated with the SADA, to which DFID is providing institutional support in order to create an effective vehicle for

facilitating and coordinating just this sort of development initiative. A separate Business Case is being developed in parallel for funding of the MV site in northern Ghana. However, given the innovative nature of the approach, and the high-profile debate that has surrounded it,[xvi] DFID has agreed with the MV Project that any support to a new MV would be accompanied by rigorous independent evaluation of the approach.

11.5. The aim of the independent evaluation is to strengthen the evidence base around MV interventions to inform decisions on possible scaling up, and to assess value for money from the use of UK taxpayer resources. The objectives of the evaluation are therefore to use rigorous and credible methods to:

a. Estimate the impact of the MV package of interventions within a cluster in northern Ghana over a 10-year period, reporting at regular intervals as data become available; and
b. Assess its cost-effectiveness compared with credible alternative uses of the resources.

11.6. The MV project team has produced a detailed discussion paper on how evaluation could be conducted of the proposed MV in northern Ghana.[xvii] It notes that random sampling across a set of MV sites and control sites is not possible, given that this is effectively a single community-level intervention, with interventions delivered across the cluster, so that it is difficult to split part of the MV site to assess various interventions. However, the paper notes that key evaluation questions can be answered by employing a mix of evaluation methods, including:

- longitudinal household-level assessments over time;
- periodic assessment of impacts against interventions;
- non-randomised 'plausibility' evaluation against a separate local matched comparison group;
- comparison against a separate intervention such as cash transfers;
- comparison against regional trends;
- "stepped-wedge" assessment of interventions introduced sequentially within parts of the MV.

11.7. These various options, and others, will be considered during the initial determination of feasibility and key design features of the independent evaluation.

12. Competition Criteria

12.1. The consultants need to demonstrate proven experience in working on monitoring and impact assessment comparable to that of rural Ghana, including fieldwork. They need to demonstrate a thorough grasp of the issues and present realistic monitoring and evaluation solutions directly related to the MV in northern Ghana.

12.2. Bids will be reviewed according to the following criteria (and weightings):

- Quality of Personnel (including, but not limited to, appropriate seniority/expertise, appropriate mix of skills, contacts/networks) **(30%)**

- Evidence of capacity to undertake work as set out in TORs **(20%)**

- Methodology (including use/numbers of days input) to develop cost-efficient innovative solutions to answer the evaluation questions **(25%)**

- Commercial **(25%)**

13. Performance Requirements

13.1. The impact of the project will be better informed, evidence-based decision-making that increases the effectiveness of future development interventions, based on improved understanding of the effectiveness of the MV model and integrated rural development approaches, and how to evaluate them. The success of the project will be determined by progress against the logical framework (included in the annex pack).

14. Format and Content of Responses

14.1. Bid responses should not exceed 50 pages (size 12 font, single-spaced lines), excluding CVs and other annexes. There is no obligation for evaluators of the bids to read the latter.

14.2. The Invitation to Tender documentation contains full guidance for suppliers. Suppliers must raise any questions relating to the TORs using the process for tender clarification set out.

22 September 2011

APPENDIX B. BALANCING TESTS

Community-level variables

	MV	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Hand dug wells (number)	6.9	0.5	0.751	0.2	0.901	0.8	0.674	0.5	0.779
Borehole wells (number)	2.6	0.3	0.591	0.2	0.801	0.5	0.502	0.3	0.668
Primary school %	86.7	6.9	0.271	8.3	0.250	5.4	0.455	-2.9	0.650
JHS %	65.7	7.5	0.429	7.5	0.496	7.5	0.496	0.0	0.999
SHS %	22.9	-1.3*	0.079	-1.4*	0.092	-1.1	0.180	2.9	0.689
Health centre %	25.7	-14.1*	0.069	-22.9**	0.010	-5.3	0.549	17.6**	0.025
CHPS %	40.0	8.8	0.306	11.8	0.241	5.9	0.556	-5.9	0.566
Market %	51.4	-3.1	0.772	-4.5	0.721	-1.6	0.897	2.9**	0.008
Distance nearest market (Km)	7.1	2.5	0.170	3.9*	0.065	1.1	0.594	-2.8	0.226
Population with health insurance %	47.8	16.7***	0.000	7.1	0.260	26.1***	0.000	19.1***	0.000
Children vaccinated against measles %	80.8	5.2	0.216	1.2	0.807	9.2*	0.057	8.1*	0.063
Farming main activity %	100.0	-1.5	0.457	0.0	0.981	-3.0	0.207	2.9	0.317
Maize main crop %	45.7	-16.2**	0.020	-14.7*	0.069	-17.8**	0.029	-3.0	0.637
Millet main crop %	51.4	12.9**	0.039	14.5**	0.047	11.4	0.116	-3.0	0.565
Shea butter main non-agricultural activity %	87.5	-4.4	0.583	-7.4	0.433	-1.4	0.881	6.6	0.511
Extension officer visits the community %	62.9	-12.6	0.226	-21.5*	0.075	-3.8	0.750	17.6	0.152
Cooperative %	28.6	-22.6**	0.002	-28.6**	0.001	-16.4**	0.043	12.1**	0.018
NPK fertiliser price (Cedis)	41.7	-0.4	0.707	-1.8	0.135	0.9	0.421	2.7**	0.017
Male wage (Cedis)	4.8	-0.1	0.750	-0.4	0.101	0.3	0.281	0.6**	0.013
Cow price (Cedis)	673.4	-6.8	0.880	-45.1	0.423	31.4	0.577	76.5	0.145
Goat price (Cedis)	63.7	8.4*	0.078	8.3	0.135	8.6	0.122	0.3	0.961
Guinea fowl price (Cedis)	12.2	0.1	0.904	-0.6	0.461	0.7	0.344	1.3	0.135
Chicken price (Cedis)	10.9	-1.5**	0.020	-1.6**	0.032	-1.4*	0.064	0.2	0.772
Gari price (Cedis)	2.4	0.3	0.206	0.5	0.104	0.2	0.515	-0.3	0.396
Rice price (Cedis)	4.4	-0.3	0.425	-0.4	0.446	-0.3	0.538	0.1	0.805
Major drought (% affected)	60.0	-0.3	0.971	-0.3	0.975	-0.3	0.975	0.0	0.999
Major floods (% affected)	11.4	10.9	0.172	9.5	0.308	12.4	0.182	2.9	0.769
CBO/NGO (% presence)	25.7	-6.1	0.470	-12.9	0.186	1.7	0.861	15.5	0.138

	MV	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Women's Group (% presence)	48.6	26.7**	0.006	19.3*	0.078	34.0**	0.002	14.7	0.155
Parent-Teacher Assoc. (% presence)	51.4	13.2	0.169	11.4	0.299	15.9	0.151	4.5	0.684

Household and individual level variables (community-level averages)

	MV average	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Demographics									
Household size	6.9	-0.13	0.692	-0.44	0.249	0.18	0.641	0.62	0.127
Female-headed household %	0.12	0.03	0.412	0.02	0.660	0.04	0.331	0.02	0.571
Polygamous	0.20	-0.04	0.110	-0.06**	0.038	-0.02	0.487	0.04	0.119
Migration									
Number of in-migrants per HH	0.37	-0.14*	0.073	-0.20**	0.029	-0.08	0.360	0.11	0.231
Number of out-migrants per HH	0.07	-0.01	0.785	0.02	0.733	0.01	0.794	0.00	0.946
Education									
Average years of schooling	1.8	0.25	0.228	0.01	0.974	0.41*	0.086	0.41	0.119
NAR primary	0.60	0.09*	0.062	0.08	0.154	0.10*	0.069	0.02	0.697
NAR JHS	0.09	0.05	0.110	0.04	0.317	0.07*	0.074	0.03	0.465
NAR SHS	0.05	0.01	0.768	0.00	0.865	0.01	0.736	0.00	0.809
Time use in the household									
Fetching wood (minutes)	152	0.62**	0.038	0.50	0.152	0.75**	0.032	0.25	0.544
Collecting water (minutes)	175	0.22	0.396	0.15	0.619	0.29	0.335	0.14	0.660
Shocks									
Affected by drought %	0.77	0.03	0.503	0.06	0.300	-0.01	0.930	-0.07	0.302
Affected by floods %	0.57	0.01	0.919	0.10	0.130	-0.09	0.182	-0.19**	0.008
Water and sanitation									
Improved water %	0.69	0.03	0.603	0.02	0.765	0.04	0.551	0.02	0.771
Improved sanitation facility %	0.08	-0.01	0.632	0.00	0.954	-0.03	0.443	-0.02	0.450
Assets									
Asset index (0,1)	2.26	-0.14	0.112	-0.15	0.148	-0.13	0.197	0.02	0.889
Credit and savings									
Member of *susu* %	0.16	-0.07**	0.013	-0.06*	0.070	-0.08**	0.013	-0.02	0.494
Any loan over last year %	0.06	-0.03**	0.046	-0.03*	0.057	-0.03	0.124	0.01	0.617
Land									
Land owned (hectares)	4.2	-0.44	0.325	-0.34	0.515	-0.54	0.302	-0.18	0.755

Household and individual level variables (community-level averages)

	MV average	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Food security									
Not enough food in any month over last year %	0.86	0.01	0.764	-0.03	0.462	0.05	0.209	0.08*	0.068
Expenditure									
Per capita expenditure (standardised)	0.94	0.05	0.586	0.00	0.963	0.09	0.373	0.08	0.448
Poverty headcount	0.75	-0.01	0.844	0.02	0.751	-0.03	0.511	-0.05	0.346
Expenditure share of own produced food	0.51	0.00	0.951	-0.01	0.702	0.02	0.625	0.03	0.400
Employment									
Employment rate % (age 15 to 59)	0.78	-0.03	0.265	-0.02	0.613	-0.05	0.156	-0.03	0.401
Farmers %	0.90	0.04*	0.080	0.05	0.111	0.04	0.157	0.00	0.870
Doing paid work %	0.02	0.00	0.916	0.00	0.785	0.00	0.650	0.01	0.504
Owning a microenterprise %	0.19	-0.02	0.532	-0.07	0.122	0.02	0.645	0.08**	0.042
Income									
Agricultural profits (standardised)	0.34	-0.09	0.15	-0.07	0.118	-0.12**	0.008	-0.05	0.220
Marketed surplus %	0.21	-0.01	0.600	-0.03	0.264	0.01	0.835	0.04	0.194
Social networks									
Any important people living elsewhere? %	0.74	0.09**	0.034	0.06	0.246	0.13**	0.012	0.07	0.199
Asked for any help over last 12 months? %	0.44	0.06	0.255	-0.06	0.331	0.18**	0.003	0.24***	0.000
Provided any help over last 12 months? %	0.46	0.04	0.479	-0.08	0.199	0.16	0.011	0.25***	0.001
Mosquito nets									
Household has a mosquito net %	0.86	0.07*	0.055	0.10**	0.009	0.03	0.458	-0.07**	0.019
Anaemia									
Haemoglobin	10.1	-0.84***	0.000	-0.89***	0.000	-0.78**	0.001	0.12	0.601
Mild anaemia %	0.72	0.13**	0.003	0.12**	0.016	0.14**	0.006	0.02	0.710

Household and individual level variables (community-level averages)

	MV average	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Malaria (children under 5)									
Malaria incidence %	-0.24	0.02	0.730	0.05	0.455	-0.01	0.853	-0.06	0.402
Nutrition (children 6 to 59 months)									
Height-for-age (Z-score)	-1.28	0.20	0.146	0.22	0.183	0.19	0.236	-0.03	0.874
Weight-for-age (Z-score)	-0.85	0.17	0.231	0.21	0.201	0.13	0.426	-0.08	0.656
Weight-for-height (Z-score)	-0.25	0.05	0.583	0.09	0.379	0.01	0.940	-0.08	0.442
Family planning and child care									
Using any birth control method %	0.10	0.02	0.418	0.00	0.961	0.04	0.182	0.04	0.188
Visited a health facility for own care or children %	0.36	0.06	0.207	-0.02	0.692	0.14**	0.003	0.16**	0.002
Child health									
Diarrhoea last 2 weeks %	0.20	0.02	0.539	0.07*	0.089	-0.03	0.408	-0.10**	0.010
Fever last 2 weeks %	0.28	0.01	0.691	0.04	0.350	-0.03	0.515	-0.06	0.144
Standardised test scores									
Raven's matrices	1.92	0.04	0.677	-0.01	0.906	0.12	0.259	0.14	0.281
Forward digit span	1.44	0.07	0.535	0.02	0.881	0.15	0.224	0.13	0.309
Backward digit span	0.93	0.05	0.613	0.00	0.985	0.13	0.249	0.13	0.293
Easy Maths	2.38	0.08	0.475	0.10	0.429	0.09	0.461	-0.01	0.947
Easy English	1.78	0.35**	0.006	0.39**	0.010	0.37**	0.013	-0.01	0.930
Advanced Maths	1.84	0.21	0.235	0.28	0.180	0.23	0.269	-0.05	0.764
Advanced English	2.17	0.08	0.555	0.07	0.634	0.13	0.375	0.06	0.655
Expected wages (parents)									
Wage primary (Cedis per day)	4.8	2.00***	0.001	1.77**	0.007	2.44***	0.000	0.67	0.402
Survival expectations									
Up to age 80	0.73	-0.01	0.644	-0.03	0.429	0.00	0.981	0.03	0.463
Time discount rates									
1-month horizon	0.11	-0.02	0.293	-0.04**	0.039	0.00	0.997	0.04**	0.042

Household and individual level variables (household-level averages)

	MV average	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Demographics									
Household size	7.2	-0.19	0.502	-0.46	0.120	0.07	0.847	0.53	0.127
Female-headed household %	0.09	0.02	0.175	0.02	0.404	0.03	0.134	0.01	0.650
Polygamous	0.22	-0.02	0.599	-0.04	0.231	0.01	0.779	0.05*	0.090
Migration									
Number of in-migrants per HH	0.10	-0.03	0.270	-0.03	0.353	-0.03	0.382	0.00	0.941
Number of out-migrants per HH	0.46	-0.16*	0.058	-0.18*	0.042	-0.14	0.189	0.03	0.737
Education									
Average years of schooling	1.9	0.06	0.767	-0.16	0.476	0.28	0.237	0.44**	0.024
NAR primary	0.61	0.08*	0.085	0.05	0.445	0.12**	0.007	0.07	0.175
NAR JHS	0.10	0.06**	0.028	0.04	0.183	0.07**	0.034	0.03	0.349
NAR SHS	0.05	0.02	0.259	0.02	0.251	0.01	0.458	-0.01	0.590
Time use in the household									
Fetching wood (minutes)	170	41*	0.058	27	0.317	56*	0.059	29	0.432
Collecting water (minutes)	182	4	0.868	-9	0.682	17	0.614	26	0.424
Shocks									
Affected by drought %	0.76	0.07*	0.090	0.11**	0.021	0.03	0.577	-0.08	0.147
Affected by floods %	0.57	-0.02	0.640	0.06	0.346	-0.11*	0.083	-0.17**	0.019
Water and sanitation									
Improved water %	0.73	-0.01	0.855	-0.03	0.708	0.01	0.919	0.03	0.701
Improved sanitation facility %	0.10	0.00	0.943	0.02	0.534	-0.02	0.507	-0.04	0.260
Assets									
Asset index (0,1)	2.4	-0.11	0.229	-0.07	0.523	-0.15	0.171	-0.07	0.596
Credit and savings									
Member of *susu* %	0.15	-0.07**	0.001	-0.06**	0.007	-0.07**	0.003	-0.01	0.786
Any loan over last year %	0.05	-0.01	0.521	-0.01	0.521	-0.02	0.215	-0.01	0.603
Land									
Land owned (hectares)	4.8	-0.28	0.695	-4.8	0.441	-0.10	0.923	0.39	0.754

Household and individual level variables (household-level averages)

	MV average	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Food security									
Not enough food in any month over last year %	0.82	0.02	0.404	-0.02	0.640	0.07**	0.040	0.09**	0.047
Expenditure									
Per capita expenditure (standardised)	1.02	0.01	0.862	0.05	0.581	-0.02	0.781	-0.07	0.522
Poverty headcount	74.0	-0.02	0.524	-0.03	0.458	-0.01	0.831	0.02	0.653
Expenditure share of own produced food	51.4	-0.02	0.581	-0.04	0.240	0.01	0.782	0.05	0.170
Employment									
Employment rate % (age 15 to 59)	0.78	0.01	0.797	0.01	0.582	0.00	0.889	-0.02	0.549
Farmers %	0.91	0.04**	0.040	0.03	0.241	0.05**	0.007	0.03	0.164
Doing paid work %	0.03	-0.01	0.292	-0.01	0.455	-0.01	0.261	0.00	0.705
Owning a microenterprise %	0.20	-0.02	0.434	-0.06**	0.034	0.01	0.733	0.007	0.123
Income									
Agricultural profits (standardised)	0.35	-0.06	0.241	0.00	0.949	-0.12**	0.003	-0.13**	0.040
Marketed surplus %	0.22	0.03	0.325	0.02	0.497	0.03	0.281	0.00	0.905
Social networks									
Any important people living elsewhere? %	0.76	0.08**	0.044	0.02	0.681	0.13***	0.000	0.11**	0.026
Asked for any help over last 12 months? %	0.45	0.01	0.891	-0.06	0.323	0.08	0.157	0.14*	0.062
Provided any help over last 12 months? %	0.53	-0.03	0.497	-0.11*	0.065	0.06	0.241	0.17**	0.013
Mosquito nets									
Household has a mosquito net %	0.81	0.09**	0.009	0.13***	0.000	0.04	0.272	-0.09**	0.017
Anaemia									
Haemoglobin	10.0	-0.47**	0.003	-0.55**	0.002	-0.44**	0.017	0.10	0.571
Mild anaemia %	0.74	0.10**	0.004	0.10**	0.005	0.09**	0.031	-0.01	0.793

Household and individual level variables (household-level averages)

	MV average	CV-MV	P-value	CVN-MV	P-value	CVF-MV	P-value	CVF-CVN	P-value
Malaria (children under 5)									
Malaria incidence %	0.22	0.01	0.764	0.01	0.787	0.01	0.818	0.00	0.961
Nutrition (children 6 to 59 months)									
Height-for-age (Z-score)	-1.29	0.05	0.623	0.08	0.497	0.03	0.841	-0.03	0.828
Weight-for-age (Z-score)	-0.89	0.03	0.745	0.10	0.378	-0.03	0.725	-0.12	0.325
Weight-for-height (Z-score)	-0.29	0.06	0.420	0.13	0.146	0.00	0.964	-0.12	0.126
Family planning and child care									
Using any birth control method %	0.10	0.01	0.698	0.00	0.909	0.02	0.510	0.02	0.368
Visited a health facility for own care or children %	0.35	0.04	0.114	-0.01	0.808	0.12**	0.001	0.13**	0.002
Child health									
Diarrhoea last 2 weeks %	0.18	0.05*	0.055	0.07**	0.011	0.03	0.300	-0.04	0.191
Fever last 2 weeks %	0.27	0.02	0.393	0.04	0.266	0.00	0.963	-0.04	0.283
Standardised test scores									
Raven's matrices	1.91	0.03	0.781	-0.03	0.813	0.09	0.464	0.12	0.354
Forward digit span	1.48	0.01	0.916	-0.09	0.370	0.10	0.235	0.19*	0.056
Backward digit span	0.92	-0.03	0.764	-0.10	0.368	0.04	0.655	0.14	0.149
Easy Maths	2.40	0.05	0.549	0.04	0.715	0.06	0.501	0.03	0.725
Easy English	1.65	0.15*	0.089	0.18*	0.094	0.14	0.191	-0.05	0.665
Advanced Maths	1.90	0.14	0.211	0.09	0.500	0.19	0.147	0.10	0.407
Advanced English	2.08	0.15	0.103	0.14	0.212	0.17	0.151	0.03	0.835
Expected wages (parents)									
Wage primary (Cedis per day)	4.94	1.9***	0.000	1.7**	0.016	2.2***	0.000	0.4	0.629
Survival expectations									
Up to age 80	0.73	-0.03	0.304	-0.04	0.181	-0.1	0.755	0.03	0.299
Time discount rates									
1-month horizon	0.10	-0.02	0.126	-0.03**	0.023	-0.1	0.513	0.02	0.127

[i] Millennium Promise (2010), "Harvests of Development in Rural Africa: The Millennium Villages After Three Years."
[ii] For example, Michael Clemens and Gabriel Demombynes (November 2010), "When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages," World Bank Policy Research Working Paper 5477.
[ii] See for instance, DFID (2011), "UK Aid: Changing lives, delivering results."
[iii] Ghana Living Standard Surveys (GLSS) 3 and 5, conducted in 1992 and 2006.
[iv] World Bank staff calculations, based on GLSS5 in 2006.
[v] CEPA and ODI (October 2005), "Economic Growth in Northern Ghana," for DFID.
[vi] World Bank (March 2010), "Tackling Poverty in the Northern Ghana."
[vii] Jeffrey Sachs (2005), "The End of Poverty: Economic Policies For Our Time."
[viii] Government of" Ghana (2010), SADA Act Number 805.
[ix] SADA (2010), "SADA Strategy and Work Plan 2010 - 2013: A Sustainable Development Initiative for the Northern Savannah", Savannah Accelerated Development Authority.
[x] www.millenniumvillages.org
[xi] The first Millennium Village was launched (in Sauri, Kenya) in 2004 and the next (Koraro, Ethiopia) the following year. Both these sites were expanded to current scale and the other sires launched in 2006.
[xii] Millennium Promise (2010), "Harvests of Development in Rural Africa: The Millennium Villages After Three Years."
[xiii] Michael Clemens and Gabriel Demombynes (November 2010), "When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages," World Bank Policy Research Working Paper 5477.
[xiv] See for instance, DFID (2011), "UK Aid: Changing lives, delivering results."
[xv] DFID (April 2011), "DFID-Ghana Operational Plan 2011 – 2015."
[xvi] The debate has been conducted primarily on the blog sites of the Center for Global Development (http://www.cgdev.org/section/opinions/blogs), Millennium Promise (http://blogs.millenniumpromise.org/) and AidWatch (http://aidwatchers.com/tag/millennium-villages-project/).
[xvii] The Earth Institute, Columbia University, and Millennium Promise (January 2011), "The Savannah Accelerated Development Authority – accelerating and sustaining development in Northern Ghana: Monitoring and Evaluation discussion paper."