UK Department for International Development

Impact Evaluation of a New Millennium Village in Northern Ghana

INITIAL DESIGN DOCUMENT

Final Version, 25th March 2013

ITAD

In association with

Participatory Development Associates Ltd.

IDS Institute of Development Studies

London School of Hygiene & Tropical Medicine

For more information about this project, visit
http://www.ids.ac.uk/project/millennium-villages-in-northern-ghana-impact-evaluation

Initial Design Document

PO 5603 – Millennium Village Impact Evaluation, Ghana

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3ie	International Initiative for Impact Evaluation
ANCOVA	Analysis of covariance
CBA	Cost Benefit Analysis
CCT	Conditional Cash Transfer
CE	Cost effectiveness
CEA	Cost effectiveness analysis
CHPS	Community-based Health Planning and Services
CIDA	Canadian International Development Agency
CSPro	Census and Survey Processing System
DALY	Disability-adjusted Life Years
DD	Difference in Difference
DHS	Demographic and Health Surveys
EAG	Evaluation Advisory Group
EI	Earth Institute
FGD	Focus group discussion
GHC	Ghanaian Cedi
GoG	Government of Ghana
GSS	Ghana Statistical Service
HAZ	Height-for-age Z-score
HH	Household
ICC	Intra-class Correlation Coefficient
ICT	Information Communication Technologies
IDD	Initial Design Document
IPA	Innovations for Poverty Action
ISSER	Institute of Statistical, Social and Economic Research (go back and find first instance)
J-Pal	Jamal Poverty Action Lab
JHS	Junior High School
JSS	Junior Secondary School
LEAP	Livelihood Empowerment Against Poverty Programme
M&E	Monitoring and evaluation
MDGs	Millennium Development Goals
MPA	Millennium Promise Alliance
MPI	Multidimensional poverty index
MV	Millennium Village
MVP	Millennium Village Project
NORPREP	Northern Region Poverty Reduction Programme
OLS	Ordinary Least Squares
OPHI	Oxford Poverty and Human Development Index
PDA	Participatory Development Associates Ltd
PE	Process Evaluation, MVP
PPVA	Participatory Poverty and Vulnerability Assessment

PRA	Participatory Rural Appraisal
PRG	Peer Review Group
PVA	Poverty and vulnerability assessment
PWR	Participatory Wealth Ranking
QA	Quality Assurance
QALY	Quality-adjusted Life Year
SADA	Savannah Accelerated Development Authority
SUTVA	Stable unit of treatment assumption
ToR	Terms of Reference
U5MO	Under-five mortality
USAID	United States Agency for International Development
VSLA	Village Saving and Loan Associations
WHO	World Health Organisation
WCA	West and Central Africa
VfM	Value for Money

Executive Summary

E1. The Millennium Villages Project (MVP) has been designed to demonstrate how an integrated approach to community-led development can translate the international Millennium Development Goals (MDGs) into results. The MVP approach has been piloted in Kenya and Ethiopia and in 2006, launched at scale to reach nearly half a million people across 10 countries throughout sub-Sahara Africa. The UK Department for International Development (DFID) has agreed to provide a grant of £11.5 million to implement a new Millennium Village in northern Ghana. The new MVP will run for five years from 2012 to 2016 with interventions targeting a cluster of communities with a total population of around 30,000 people.

E2. While MVPs across Africa have established their own monitoring and evaluation systems, there has also been a noticeable gap in the evidence of overall effectiveness and impact. Recent reports have used before-and-after analysis within the MV sites, leading to criticisms of the extent to which results can really be attributed to the MVP, and a lack of independent rigorous evaluation. DFID has agreed that the new MVP in northern Ghana will be accompanied by an independent evaluation. The evaluation is to be governed by an Evaluation Advisory Group that provides advisory support from a range of stakeholders, and a Peer Review Group (PRG, managed by the International Initiative for Impact Evaluation (3ie)) to review the technical and scientific quality of the evaluation. The evaluation will build on, expand, and validate the MVP's own monitoring and evaluation of the MV site.

E3. This Initial Design Document sets out the conceptual approach and methodological design for the independent impact evaluation. The evaluation will use a difference-in-difference (DD) approach, 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. DD allows the evaluation to isolate the MVP impact on the outcomes (including poverty, child development, under-nutrition, and child mortality), from effects of other variables changing over time. A randomised control trial was not feasible in this case as the methodology would have required a change in the MV model – a model that is based on the implementation of an *integrated package* of interventions across a single site. 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.

E4. In addition to measuring the impact (net effect) of the MVP, the evaluation will assess:

- *Synergistic effects:* The evaluation will estimate the synergistic effect of the intervention, as generated by the simultaneous investments in all sectors of the village economy. The implication of the MVP approach is that it can generate outcomes in excess of the outcomes produced by the sum of the single project components. This should mean that the MVP is expected to be less expensive than comparable interventions achieving the same unit outcomes (such as one life saved). This will be estimated as part of the cost-effectiveness analysis.
- *Externalities:* It is likely that the MVP will generate three types of externality. First, benefits may spread from the MVP villages to the nearby villages. For example, nearby localities may access services provided by the project, or reductions in infection risk in the MVP areas may benefit non-project villages. These effects will be detected through the stratification of the sample of control communities by distance; thereby building in an additional treatment arm so that the control consists of both 'faraway' and 'nearby' communities. Second, MVP may lead to changes in expenditure allocations and decisions by district officials and other project expenditure (such as by non-government organisations). This will be detected through two additional modules: one collecting data on projects by any funding body in the locality, with the other module collecting data on district assembly projects and expenditure in project and control localities. Third, there is a possible 'demonstration' effect whereby other districts adopt similar policy packages to those of the MV districts (e.g. the free distribution of bednets, fertilisers, etc.). This will be documented through the collation of district budget and expenditure data, and through qualitative institutional assessments.

- *Sustainability:* The extent to which the effects continue after the project implementation will be assessed in two main ways. First, the evaluation will re-survey the project and control villages after 10 years (i.e. five years after the project intervention has ended). Second, the evaluation plans to capture irreversible welfare indicators which are known to be good predictors of long-term welfare. This will include the cognitive abilities of children, reductions in stunting prevalence, and the dynamics of asset accumulation.
- *Cost-effectiveness:* The evaluation will explore whether MVP is the most cost-effective way to achieve comparable results, by undertaking three main types of assessment. First, by conducting a cost-effectiveness analysis of each component of the programme – with the expectation that the synergistic impact of MVP will result in each component of MVP being more efficient than similar programmes carried out separately. Second, a feasibility assessment of whether a cost-benefit analysis is possible for the entire MVP, or whether the monetisation of outcomes is more clearly suited for individual components. And third, a multi-dimensional poverty measure as an alternative to a monetised value of development. The possibility of simulation for the scaled-up economy-wide impact, along with some attempts at local area economy-wide impact, will also be considered.
- *The heterogeneity of impact:* This will be investigated along geographic and household characteristics. First, a disaggregated analysis of impact across the districts is required, with the sample of control villages stratified in such a way to allow this type of analysis. Then, two other geographic distributions will be considered across the districts. The first distinguishes between remote and centric villages with respect to main markets and service facilities. The second separates a core and periphery of the project area – with the geographic centre assumed to benefit to a greater extent. Household characteristics include gender, but also landholdings and education of the household head. The impact of the intervention will be disaggregated by subgroups or by quintiles of the relevant characteristic. Non-parametric, semi-parametric and switching regression models will be employed in the analysis.

E5. The evaluation will draw upon data from three full-length survey rounds in 2012 (baseline), 2014 and 2016 using an adaptation of the MVP Household questionnaire with additional modules on income, expenditure, in/out migration and social networks. Alongside this instrument, is an Adult (female, male) questionnaire based on the Demographic Health Survey (DHS) instrument there is a Facilities questionnaire (to capture characteristics of health and education facilities), a Village questionnaire (to capture data on land area, distance to facilities, economic activities, market prices, shocks and development projects), an anthropometric module measuring height and weight of children under five, and blood tests for the measurement of haemoglobin and anaemia. Additional data modules will be conducted on education and cognitive tests for school-age children, plus on expectations and time preferences – in order to assess expectations of survival, incomes and educational returns, as well as attitudes to risk. In 2013 and 2015 a shorter version of the Household questionnaire will be conducted to track progress for key poverty reduction variables (i.e. around the expenditure module, together with some core MDG outcomes like school attendance). The addition of these two extra 'mini' rounds allows some variables to be measured on annual basis.

E6. In addition to quantifying the net effect of MVP, the impact evaluation also includes supporting qualitative modules that aim to better understand how and why change has occurred. There are four key qualitative methods. First, a Poverty and Vulnerability Assessment (PVA) that will track local and multi-dimensional perspectives of wealth and wellbeing. Second, an Interpretational Lens that will take preliminary quantitative findings from the survey data and obtain local feedback and interpretation around emerging themes of the analysis. Third, an Institutional Assessment that will capture empowerment and institutional change, particularly between community, district and regional levels. And fourth, a number of Reality Checks that provide mini-anthropological studies to better understand how MVP affects the realities of people, and any unintended consequences.

E7. The integration of survey data with the qualitative modules will be achieved in a number of ways. Primarily it will be achieved through the Interpretational Lens approach, where the sequencing of this part of the

qualitative module will enable feedback and local perspectives to be gathered about the emerging (and statistically representative) findings. Secondly, the PVA will take place in a sub-sample of matched communities (treatment, and control) with direct links made between the quantitative household survey data and households taking part in the qualitative modules. Third, the quantitative and qualitative experts will work together to develop theories of change (based on the available literature). These will be used to provide working hypotheses that will guide the analysis across the different datasets, inform any adjustments to the quantitative survey instruments, and provide a focus for further qualitative investigation.

E8. The evaluation will produce a number of reports that will be peer reviewed by the PRG. These will include the Baseline (2013), Mid-term (2014), and Final evaluation reports (2016), plus a number of working papers and journal articles.

1 Introduction

1. This document is the Initial Design Document (IDD) for the Impact Evaluation of a new Millennium Village in Northern Ghana, and sets out our conceptual approach and methodological design.[1] The Millennium Village (MV) model provides an integrated package of interventions that aims to lift rural communities out of poverty. The central hypothesis is that by addressing the most immediate capital deficiencies in communities and households through a form of local 'big push', this provides the necessary conditions for reaching the threshold required to move towards local resilience and self-sustaining economic growth. A key part of the approach is to improve agricultural productivity and market development, enabling people living in rural areas to save and accumulate wealth, stimulating investment and diversifying into non-farm work. To date, MV projects have set up comprehensive Monitoring and Evaluation (M&E) systems, which have been used to assess progress and adapt implementation. This has also enabled the reporting of results within the MV sites but because this has been based on a before-and-after analysis (Millennium Promise 2010), there has been criticism of the extent to which results are attributed to the MVs and the lack of independent rigorous evaluation (e.g. Clemens and Demombynes 2010).

2. This impact evaluation will address this gap by providing robust evidence on the effectiveness and impact of the MV approach. The evaluation is an independent evaluation covering a 10-year period,[2] and will cover the costs of expanding the MV project's own M&E surveys, conducting additional baselines and surveys in comparison sites, and applying qualitative evaluation techniques.[3] Given the level of interest and criticism of past evaluations of the MV approach, it is particularly important that this impact evaluation stands up to external scrutiny by adhering to the following principles:[4]

   - **Independence:** The evaluation is independent in that it is led by people with no other involvement in the MV project. This is important to prevent a conflict of interest between implementation and evaluation, as well as to reduce possible research biases, either real or perceived.
   - **Transparency:** All evaluation reports will be quality assured to ensure they meet the required standard, and then published with a view to being publicly accountable, as well as to facilitate lesson learning. The reports will be made accessible to all, including key stakeholders, recipients and taxpayers (via the DFID website), as well as through the websites of ITAD and IDS.
   - **Methodology:** The impact evaluation will assess the impact (net effect) of the MV as outlined in the evaluation Terms of Reference (ToR). This will include an assessment of why and how things work checking for unintended effects, and testing the theory and assumptions behind the intervention. To do this, the impact evaluation will adhere to the internationally-agreed evaluation OECD/ DAC criteria: relevance, effectiveness, efficiency, impact and sustainability.

1.1 The Peer Review process

3. This document presents the detailed methodological design for the impact evaluation of the MV in northern Ghana. It has been submitted to an independent Peer Review Group (PRG) process in a number of stages. This staggered process has been necessary to enable the baseline data to be collected in advance of the MV implementation, which itself is constrained by practical challenges due

[1] *Terms of Reference for the Impact Evaluation of a new Millennium Village in Northern Ghana*, PO 5603 MV-Evaluation, ITT Volume 3, Department for International Development (DFID), UK. See Annex A.
[2] For contractual reasons, the evaluation team is commissioned to undertake the work for the first five years only (2012-2016 inclusive). For this reason, this design document is primarily concerned with the first five years of the evaluation.
[3] See: *Business Case: Millennium Village Evaluation in Northern Ghana*, DFID, 2011.
[4] See: "*Principles governing the relationship between DFID, ITAD and the Millennium Villages Project (MVP) for the evaluation of the Millennium Village in Northern Ghana project*", internal document, March 2012, DFID-Ghana.

to the start of the rainy season in the north. The PRG is coordinated by the International Initiative for Impact Evaluation (3ie) and consists of five independent reviewers.[5]

4. The following elements of the methodology have been submitted to and approved by the PRG:

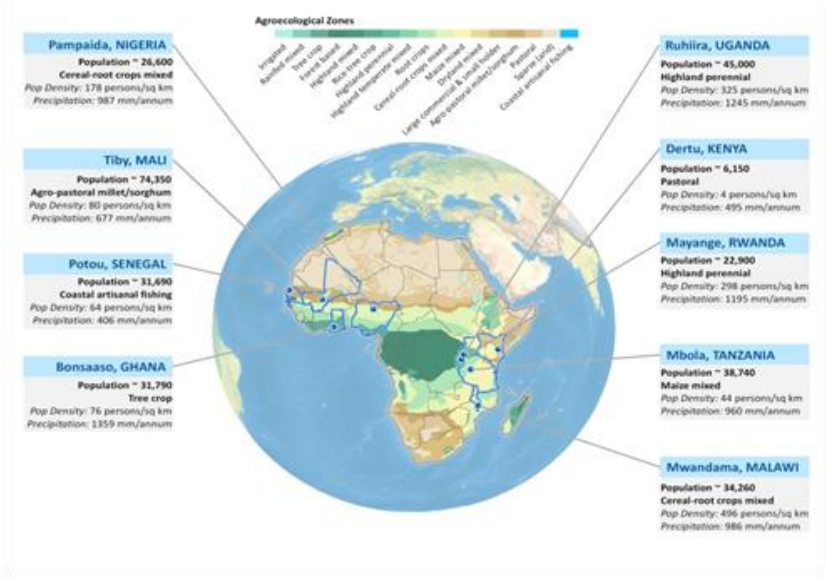
- An initial IDD was approved by the PRG on 13th April 2012. This covered the evaluation approach and key survey instruments, as well as the quality assurance processes that were necessary for the baseline data collection to commence.
- The full sampling strategy including the selection of control villages in Builsa and West Mamprusi districts was approved on 23rd May 2012.
- This document, the complete IDD, which sets out the full methodology for the impact evaluation including details of the qualitative methodologies, the approach to cost-effectiveness analysis, additional modules on education and cognitive tests, time preferences and expectations.

5. The IDD is structured in the following way: Firstly, the document starts with an overview of the MV, describing the intervention and its aims (Chapter 1). The rest of the document then provides details on the evaluation approach and methodology, including the programme theory, the quasi-experimental design, and the quantitative survey tools (Chapter 2). The cost-effectiveness methodology (Chapter 3) and the qualitative methodology (Chapter 4) are then covered, followed by the governance arrangements for the evaluation (Chapter 5). Finally, the IDD ends with a summary of our communication strategy (Chapter 6).

1.2 Programme description

6. At the UN Millennium Summit in September 2000, world leaders adopted the Millennium Declaration, committing nations to a new global partnership to reduce extreme poverty and address pressing challenges of hunger, gender inequality, illiteracy and disease. Targets were set for these Millennium Development Goals (MDGs) to be achieved by 2015. The Millennium Villages Project (MVP)[6] is explicitly linked to achieving the MDGs and creating wealth at the local level. The MVP addresses an *integrated* and *scaled-up* set of targeted investments based on the recommendations of the United National Millennium Project (UNMP). The interventions cover food production, nutrition, education, health services, roads, energy, communications, water supply and sanitation, enterprise diversification, environmental management and business development. The initiative makes use of

Figure 1: Locations of MVPs



[5] The peer reviewers are: Annette N. Brown (Deputy Director, 3ie), Thomas de Hoop (Evaluation Officer, 3ie), Robert Darko Osei (University of Ghana), Christopher Udry (Professor of Economics, Yale University), and Howard White (Executive Director, 3ie).

[6] MVP is used throughout this document to denote both the project and the project implementation team. Where appropriate the M&E unit responsible for data collection is identified separately. The M&E is led by the Earth Institute, Columbia University, USA.

community decision-making and uses science-based technologies and techniques, such as agroforestry, insecticide-treated malaria bednets, antiretroviral drugs, remote sensing, and geographic information systems. The MVP is a 10-year initiative with two five-year phases.

7. The *first phase* focuses on achieving quick wins,[7] especially in staple crop production and disease control, and on establishing basic systems for integrated rural development that help communities escape the poverty trap and achieve the MDGs. The interventions are delivered at a cost of approximately $120 per capita per year of which the MVP contributes around half to complement funds from the host government, the local community and other partners. The *quick wins* are complemented by infrastructural improvements (buildings, roads, energy, water and sanitation), which are necessary to strengthen the primary health care and education services, as well as facilitate development of agricultural markets and business development. The revitalisation and strengthening of community institutions (local government, sector based institutions, etc.) and the implementation of strategies to increase women's participation and leadership are also important facets of the MVP approach. The *second phase* focuses more intensively on commercialising the gains in agriculture and continuing to improve local service delivery systems to support the local scale up.

8. The MVP was piloted in Kenya (Sauri) and Ethiopia (Koraro) in 2005 and launched at scale in 2006 to reach nearly half a million people across 10 countries. Figure 1 shows a selection of MVP sites. The MVP is currently working in communities across 14 sites (or 'clusters') with catchment areas reaching up to 70,000 people per site. Sites are identified according to a number of criteria:

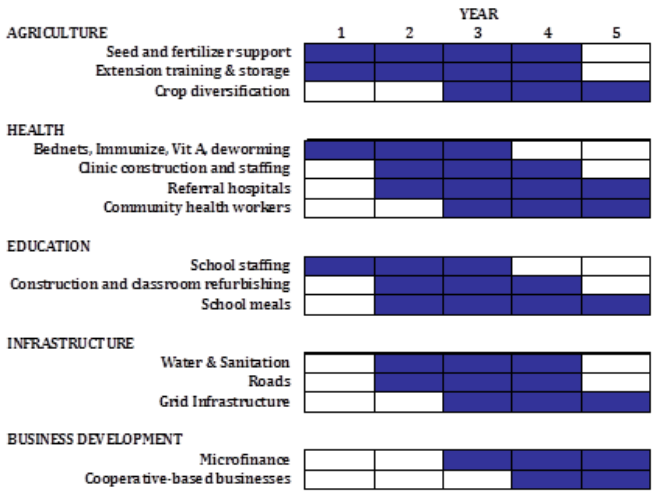
   - First, all sites were located in hunger hotspots (areas with more than 20% of children under the age of five were underweight for their age).
   - Second, the sites were selected to represent a cross- section of Africa's key agro-ecological zones and farming systems in order to test whether the initiative could generate a 'green-revolution-style' breakthrough in smallholder farm productivity.
   - Third, sites are located in countries where the national government is committed to partnering in the initiative and to the MDGs more broadly.

9. Central to the MVP approach is the value *of integrated community-based investments* in scientifically proven interventions, delivered simultaneously rather than as one-off investments. The premise is that a critical platform of basic needs must be reached before economic development can take off. The interventions are implemented in a multi-layered, multi-sectoral and integrated manner. The reasons underpinning this approach are rooted in the following explanations:

   - To address multiple objectives across a range of sectors (health, poverty, disease control, nutrition, etc.), with a holistic strategy that uses a range of tools (community-based clinics, diversified local food production, malaria control, etc.).
   - To enable multiple tools to produce synergistic gains, with each supporting a main objective but also contributing to progress on several or all of the other goals.

[7] "Quick wins" are defined as scientifically proven interventions that can achieve wide coverage and lead to dramatic gains in short periods of time. The quick win interventions are: (1) distribution of improved seeds and fertiliser; (2) distribution of long-lasting insecticide treated bed nets; (3) basic immunisations; (4) Vitamin A campaigns; and, (5) community wide de-worming to reduce levels of intestinal *parasitosis*.

Figure 2: Indicative Timeline for Core MVP



10. MV communities carry out these integrated targeted investments such as in agriculture, health, education, infrastructure and business development. Community-based assessments and participatory planning processes are used to identify the timing, sequencing and level of priority afforded to the interventions in the different sites. Figure 2 shows the generic sequence of activities for MVP interventions in the first five-year phase.

11. The guiding principle of the MVP budget framework does not however imply a top-down set of fixed interventions across every community, but rather a flexible and coherent approach through the implementation of village-level budgeting and multi-sector budgeting to help ensure that communities gain access to a basic set of goods and services.

The Millennium Village Project in Ghana

12. In Ghana, Bonsaaso was selected as an MV in 2006, and is located in the Amansie West District in the Ashanti Region. The total population of 35,000 people is spread over 389 square kilometres in six villages. A three-year review of the MVP in Bonsaaso (2006-09) was published by Millennium Promise (Mensah-Homiah et al. 2011). The report highlights a number of results against the MDG indicators, based on its own M&E datasets. This evaluation will similarly assess the results and impacts of the MV, but in an independent and more rigorous manner (see Chapter 2). The review of the Bonsaaso MV also highlights a number of lessons that are useful to consider in the design of this evaluation. These include issues concerning: the *effectiveness of delivery mechanisms* (e.g. delays in procuring seed and fertiliser inputs due to procurement difficulties, floods destroying farms, and the poor conditions of roads); *the complementarity or otherwise of different interventions* (e.g. business-related ventures were challenging to implement due to the poor state of the roads, inadequate on-farm and off-farm transportation, as well as the non-availability of budgetary funds for business development); *structural constraints in particular sectors* (such as in education, where results were potentially undermined by the low numbers of trained teachers); *the knock-on effects of particular interventions* (e.g. the need for new classrooms, etc., due to rising attendance rates caused by the MV project); and, *the effectiveness of institutional and management arrangements* (e.g. a few communities still do not pay for water which affects the smooth operation or management of these facilities). These lessons have been considered in the design of this evaluation, and particularly the methodology for the qualitative research (Chapter 4).

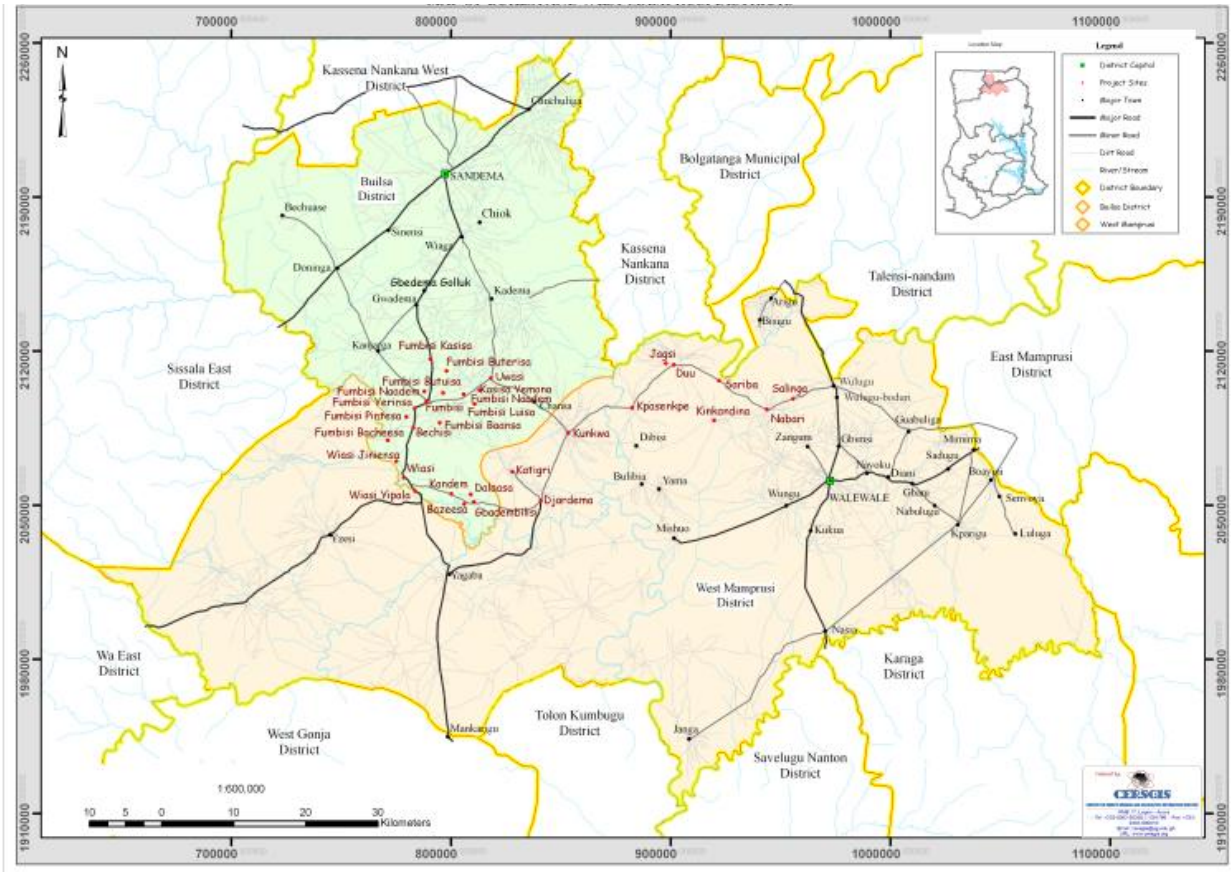
13. As part of their second phase of implementation across Africa, the Millennium Promise Alliance (MPA) approached DFID to finance a Millennium Village in rural northern Ghana. The proposal was for the design, implementation and monitoring of a five-year set of integrated interventions to accelerate development in a cluster of communities of up to 30,000 people based on the MV model, with potential to be substantially scaled up. The UK will provide £11.5 million over five years between 2012 and 2017 for the implementation of this new MV site in the West Mamprusi and Builsa districts of northern Ghana. By supporting a new MV along with an independent and rigorous evaluation, DFID and the semi-autonomous Savannah Accelerated Development Authority (SADA)[8] aim to provide evidence

[8] The SADA Strategy, 'A Sustainable Development Initiative for the Northern Savannah', emphasises mobilising and coordinating increased investment from public and private sources in order to stimulate private sector-led economic growth, along with careful monitoring and evaluation of development interventions in order to maximise impact.

that can guide further development in northern Ghana, influence policy in Ghana, and inform the international debate on the effectiveness of the MV model. As mentioned previously, past evaluations of the MV model have been mostly internal, undertaken by the Earth Institute and the MVP implementation team, and therefore have not been viewed as independent. Furthermore, they have been criticised for not being rigorous due to their reliance on before-and-after datasets[9] (often showing positive trends), without a sufficiently robust control group (to enable the measurement of the net effect in the MV site).[10]

14. Development in northern Ghana lags behind the rest of the country. While the number of poor people in the southern part of Ghana has decreased in the past 20 years by 2.5 million, it has increased in the North by almost one million. Accelerating development in the north is therefore a central theme of DFID-Ghana's Operational Plan for 2011-2015. DFID has identified the need for significant and sustained investment in order to achieve the MDGs in the north and to bring it to a level of development more in line with the rest of the country. The challenge of addressing the stubbornly high rates of poverty in the north is well recognised, and is reflected in the creation of SADA by the Government of Ghana (GoG).

Figure 3: Map of Builsa and West Mamprusi Districts



15. The outputs of the project are expected to be:[11]

- Skilled staff delivering improved basic services – for instance aiming to increase the proportion of births attended by skilled health workers from around 30% to 70%

[9] For example, the Millennium Promise (2010) *Harvests of Development in Rural Africa: The Millennium Villages after Three Years.*
[10] This has been highlighted by several commentators for the MV approach to evaluation, such as Michael Clemens. See for example: "*Impact Evaluation in Aid What for, how rigorous?*", presentation for the Royal African Society and Overseas Development Institute (ODI), Center for Global Development, 3rd July 2012, London, UK.
[11] *Business Case: Millennium Village in Northern Ghana*, DFID, November 2011.

- Women, men and young people facing fewer barriers to accessing services – for instance aiming to increase the contraceptive prevalence rate from around 14% to 40%
- Better connectivity within and beyond the Millennium Village site – such as improved roads, electricity and mobile phone penetration
- Improved income earning opportunities – such as through increased yields, access to financial services, and promoting non-farming enterprise
- Strengthened local institutions and community capacity – to increase the involvement and commitment by the local community and officials, and help to secure sustainability of the impacts

16. At the time of writing, the Millennium Promise were still conducting a detailed planning exercise in northern Ghana to plan the sequence, timing and priority afforded to the different interventions for the MVP. The site selected for MV interventions lies on the border between the Northern and Upper East regions, spanning the two administrative districts of West Mamprusi and Builsa.[12] The actual MV intervention and the planned results for northern Ghana are expected to be further refined over the first year of operations, and they will take account of information obtained from the baseline surveys.

[12] This is understood to be in order to meet SADA's requirement that the MV be located in more than one region, so to represent a broader northern savannah intervention.

2 Evaluation approach and methodology

17. This Chapter sets out the details of our evaluation approach and methodology. The first part starts by setting out the programme theory behind the MVP, before moving on to explain how the impact evaluation will test this theory using a difference-in-difference (DD) approach. This is followed by details on how each of the evaluation questions will be addressed, the sampling strategy, the selection of the control sites, and an overview of the survey instruments.

18. Before turning to the programme theory, it is helpful to first set out the key evaluation questions to be addressed by this evaluation (ToR, Appendix A). These questions were derived from discussions between DFID and other key stakeholders before commissioning the evaluation. While the central question for the evaluation concerns impact, this is not just whether the MVP has met its targets against the MDGs (although important). For many observers, it may indeed be quite reasonable to expect that the MVP will lead to some impact in making progress towards the MDGs – especially given the size of the investment amongst relatively few communities (£11.5 million in northern Ghana from DFID alone). Rather, the more interesting questions for the evaluation are around measuring the extent to which these impact/outcome variables can be attributed to the MVP, plus whether these effects are sustainable and cost-effective compared to other possible alternatives. The evaluation is therefore charged with assessing whether the simultaneous and integrated investment by MV has produced effects in excess of the sum of the individual parts (*synergistic effects*); whether this has led to *externalities* beyond the immediate MV site; whether the changes are *sustainable* after the direct intervention ends; and, whether MV is the most *cost-effective* way to achieve comparable results. These concerns underpin the core questions for the evaluation, which can be summarised as:

    1. Does the MV **deliver on promises to reach the MDG**s within the MV site? (Section 2.2, pages 17-19)
    2. What **externalities or spill-over 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? (Section 2.2, pages 18-19)
    3. Are the positive impacts of the **MV sustainable after direct implementation** of the MV project has ended? (Section 2.2, pages 19-20)
    4. Is the MV intervention package **cost effective** in the results it achieves, compared with possible alternatives? (Section 2.2, page 20, with full details in Chapter 3, pages 42-51)
    5. Does the MV package **empower disadvantaged or marginalised groups** (e.g. females, the disabled, or the elderly)? (Chapter 4, pages 54-61)
    6. Does MV achieve **additional benefits arising from synergies** across implementation of an integrated package of interventions? (Section 2.2, pages 17-18)

The section that follows focuses on the programme theory and the key assumptions and challenges behind measuring the MVP approach. Behind any intervention there are theories and assumptions guiding designers and implementers of the programme. Whether explicit or not, this 'programme theory' is helpful for evaluators in identifying and measuring not only the stated goals of the intervention, but also how it will address poverty over the short and longer-term (the poverty trap).

2.1 Programme theory

19. An impact evaluation can be defined as an analysis that measures the extent to which net changes in key outcome / impact variables can be attributed to the intervention. This evaluation follows the principles of a 'theory-based impact evaluation' (White 2009). A theory-based evaluation is simply the examination of the assumptions that underlie the causal chain from activities and inputs, through to outcomes and impact (Weiss 1998). Applying a theory-based approach to an impact evaluation simply means making use of

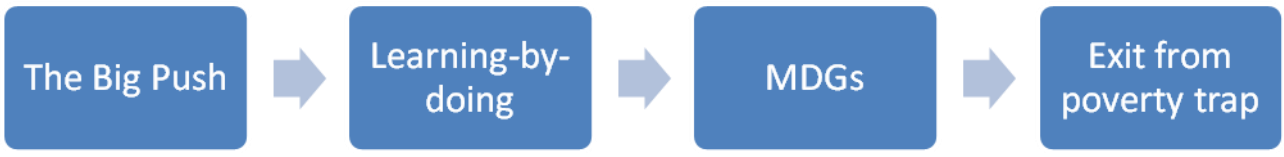
theory to uncover the assumptions and theories behind the intervention, and to test them in a robust manner. To achieve this, the MV impact evaluation combines both the use of a quasi-experimental design to test the extent of the net impact (using a treatment and control groups), with elaborations of the programme theory to help explain these findings. This first section explains in detail our understanding of the MVP's programme theory.

20. The MVP is an experimental application of the poverty trap theory presented by Jeffrey Sachs and collaborators in a paper published on the Brooking Papers on Economic Activity (Sachs et al. 2004).[13] In this paper the authors propose a revival of a classical development economic theory: at very low levels of income, the returns to capital investments are not large as assumed by mainstream economic theory. On the contrary, poor human and physical infrastructure, at low levels of incomes, result in very low returns to investments. As a consequence, extremely poor countries are 'trapped' in poverty as attempts to increase incomes have very little success. The authors then discuss at length the factors that determine poverty traps in African economies. These include: high transport costs; poor agricultural productivity; the burden of malaria and poor health; and a history of colonisation and depredation.

21. The implication of the poverty trap model is that small investments in poor countries are unlikely to bring any benefit in the long term. In order to bring countries on a sustained development path, large and simultaneous investments are required (the 'big push'). The authors then proceed to delineate the modality of how a big push could be implemented in the African continent. They find that the MDGs are the best indicators for monitoring country progress out of the poverty trap. They suggest a set of interventions that are to be implemented by central governments with the support of donors and local governments, in the following areas: agricultural productivity; health, nutrition and family planning; primary education; urban infrastructure and services; science and technology; gender equity; and regional integration. It is important to note that the authors are convinced about what interventions and best practices could be used to fast track progress towards the MDGs – drawing from the best scientific research available at the time. This point is forcefully made in the report to the UN Secretary by the Millennium Project 'Investing in Development' (Millennium Project 2005).

22. The theory exposed in the paper and the report quoted above has found a practical application in the MVP model. Here, villages (later on to become a cluster of villages with the Millennium Promise) receive a simultaneous package of interventions including: food production, nutrition, education, health services, roads, energy, communications, water supply and sanitation, enterprise diversification, environmental management and business development. Although there is scientific evidence that shows which interventions and best practices could be used to fast track progress towards the MDGs, it is acknowledged that we do not know with certainty what works in different and specific circumstances. Hence, the project requires that a process of learning-by-doing should be put in place in every community in order to select the right mix of development interventions specific to the area. The programme theory is broadly sketched in the figure below.

Figure 4: Programme Theory



23. There are a number of **assumptions** behind this theory that have important implications for the evaluation design that we will briefly discuss here:

[13] Other documents of the Millennium Promise and partners also refer to the MVP as an attempt for people to escape from the 'poverty trap' (e.g. Millennium Project 2006).

24. **First**, the idea that we know what works and that, even if we do not know what works in some specific cases, we can learn through a combination of good management, inclusive processes and good monitoring is not necessarily convincing. Many of the 'sure win' interventions suggested by Sachs and collaborators (like watershed interventions or school meals for example) have obtained conflicting evidence of success. Plus, the idea that the monitoring process set up by the intervention will be able to measure success with certainty is questionable. The risk is that even if the poverty trap model is correct, its practical implementation may fail because of the inefficacy of the specific development interventions adopted.

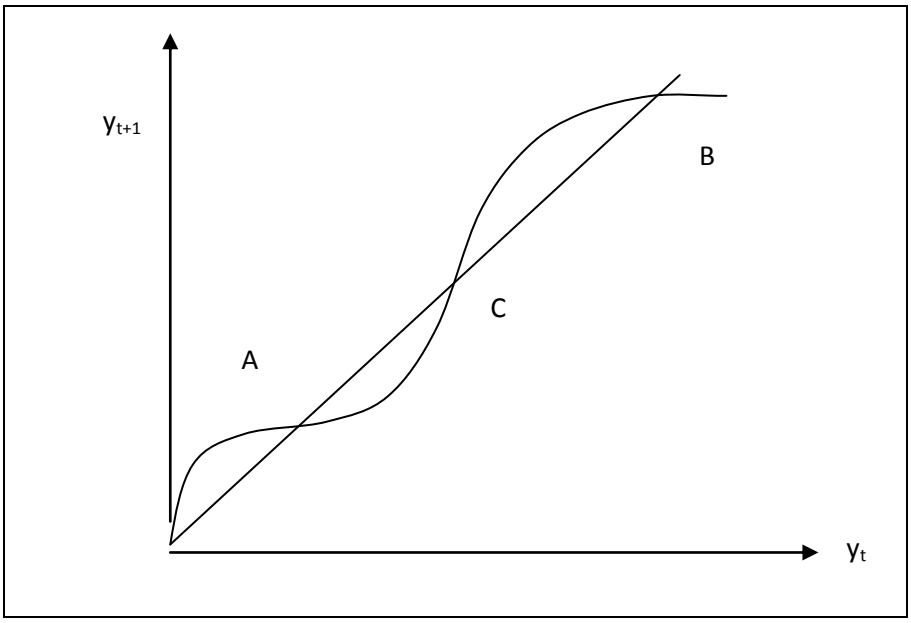
25. **Second**, it is somewhat surprising that a model built on non-linearities avoids the discussion of economies of scale effects. There are certainly synergies between intervention villages that can be realised, but due to externalities, the impact of the intervention on a particular village differs depending on whether it is implemented in only one, 10, 100 or 1,000 villages at the same time. Similarly, the impact of the intervention on a single household or individual will vary with the scale of the intervention. This is a violation of the stable unit of treatment assumption (SUTVA) which states that treatment of person *i* only affects outcomes of person *i*.

26. **Third**, the poverty trap model outlined by the project seems to neglect the rural-urban duality of the development process. Neoclassical economic theory of the dual development process predicts that the development of agricultural and rural areas will result in employment redundancy in the agricultural sector and the migration of skilled labour to urban areas. In other words, successful project villages could become smaller as skilled workers (or their children) will find more remunerative jobs in urban areas. In these circumstances comparing the performance of the project villages against the national average or other comparator group could be misleading. Certainly, we do not yet know the full dynamics of this process.

Poverty traps and the big push

27. The MV intervention is an application of the economic poverty trap model of development and an impact evaluation of the intervention is an implicit test of a 'big push' theory aimed at breaking the poverty trap. The MV intervention invests considerable resources, including government and community resources, on a wide range of activities. The recipients are localised in small communities and will be covered by the intervention for at least five years (in the case of the northern Ghana MVP).[14] Under these circumstances it is very unlikely that the project will not have an impact on a wide range of human development indicators. Indeed, it seems likely that the programme will improve incomes, reduce mortality and increase literacy. But, what is the impact size needed to claim the programme is a success?

Figure 5: The canonical poverty trap model



28. Standard neoclassical economic theory assumes that in the long term all countries converge to the income level of rich countries, while over the generations incomes of poor households converge to

[14] It has been agreed that interventions by MVP will stop after five years in the northern Ghana MVP. This is to provide a clear cut-off date for the impact evaluation, with survey data collection currently planned to continue until Year 10.

those of richer households. Much empirical evidence however is against the convergence in the rates of economic growth across countries (see for example Pritchett 1997; Ray 1998), and in economic mobility across generations (see for example Mazumder 2005; Bourguignon, Ferreira et al. 2007).

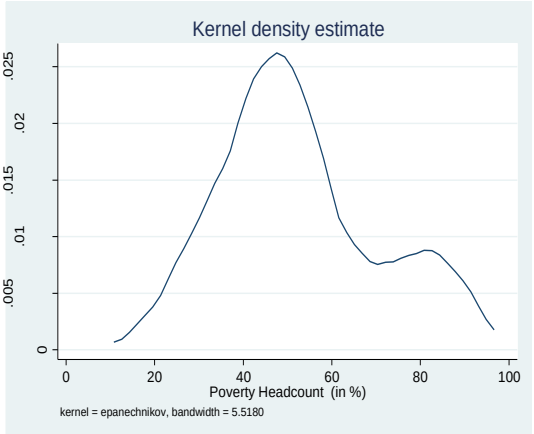
29. Poverty trap models try to explain why poverty and inequality persist among countries, generations and individuals. Generally speaking, a poverty trap is any self-reinforcing mechanism that causes poverty to persist. The fundamental intuition behind early poverty trap models is the presence of increasing returns to scale in the process of economic growth. In the canonical threshold poverty trap model (Azariadis and Stachurski 2005; Carter and Barrett 2006) poverty traps are generated by feedback mechanisms. If output depends on output scale, income increases over time as income increases, and income in the next period (y_{t+1}) is a function of income in the present period (y_t). The dynamic growth path produced by this iterative process is illustrated in Figure 5.

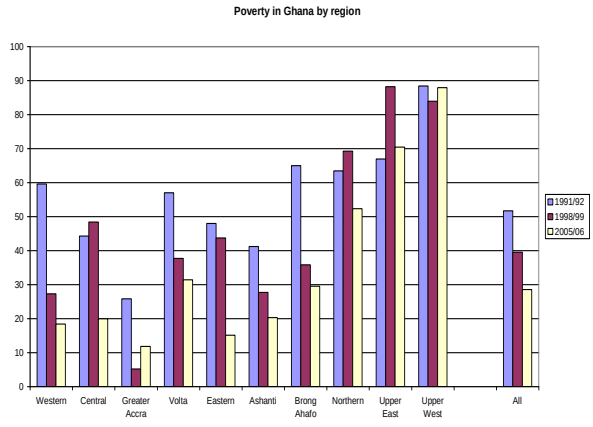
30. The 45 degrees line represents all points in which income in the current period and income in the following period are exactly equal. The figure is drawn in such a way to show three inter-temporal equilibrium points, of which two are stable (A and B) and one is unstable (C). Starting from values of income to the immediate right and left of point A, income converges to the equilibrium point A. By contrast, the equilibrium point C is unstable, and any movement from this point brings income either to point A or B. Point A is a poverty trap: small income increases beyond A leads income to converge again to A. The intersection of the phase line with the 45 degrees line at point C is the poverty trap threshold. Movements of income beyond this point push a country or household permanently from one equilibrium state to the other.

31. One implication of this model is that temporary events may have permanent consequences. Events like wars, natural disasters or wrong policies may compromise long term economic development by pushing a country below the critical threshold. In the same way a massive foreign aid programme may have long-term consequences if it is able to push the country beyond point C.

32. There are some signs that the geographic area where the programme is implemented in northern Ghana is trapped in poverty. The left chart in Figure 6 shows the distribution of poverty rates in Ghana across districts. The distribution is clearly bimodal. The pattern is highly correlated with geographic characteristics. The vast majority of districts with poverty rates around 80% are located in the north of the country (the regions of Upper West, Upper East and Northern region). More importantly, poverty trends by region (right chart) show that poverty in the northern regions has barely changed over the last 20 years while it decreased considerably in most of the remaining districts. There is district-level divergence in Ghana; the country is getting richer but the north is stuck in poverty.

Figure 6: Poverty headcounts in Ghana (based on 2000 census data and GLSS 1998/99)





33. The MV intervention effectively consists of a five-year shock[15] on a small cluster of localities. If the programme is successful, after these five years these localities will enjoy standards of living more similar to those of localities in the south of the country. By comparing project villages to other villages in the same districts over time we should observe diverging patterns of income. If the project is breaking the poverty trap, we should be able to observe incomes increasing at an increasing rate. Hence, we should observe in the project villages: a) higher incomes in any given year, and b) increasing income growth for some period.

Testing the breaking of the poverty trap

34. The most appropriate way of testing income differences between project and control villages is to apply the difference in difference (DD) estimator over the household-level series of income for the period 2012-2016.[16] This is, for example, the strategy adopted by Ravallion and Chen (2005) to assess the impact of a development programme in rural China using a household panel of income data for the period 1996-2000.

35. A formal test of 'divergence' can be conducted by looking at income growth and its determinants in regression form:

$$y_{it} = a + by_{it-1} + \boldsymbol{cZ_0} + dPT$$

36. Where y_{it} and y_{it-1} are incomes (for *i*= village or household) at current year and previous year respectively and $\boldsymbol{Z}$ is a vector of covariates which may include initial conditions in terms of household and community wealth. The variable *PT* is the interaction of project and time-specific dummies. The coefficient *b* tests the convergence hypothesis ($b<1$) and the *d* coefficients estimate the growth shift produced by the programme in each year. A similar formulation was also adopted by Jalan and Ravallion (1998) to assess the impact of an anti-poverty programme in China using household panel data for the period 1985-1990.

37. One problem in the analysis of income patterns is that the programme may not be able to sort its effect on income over the short term (five years) considered by the evaluation. In addition, an increase in income is not necessarily a result of a structural improvement in household income generating capacity. To see this consider the following expression for household income:

$$y = rA + T + S$$

38. The programme can increase income by increasing assets (*A*), the returns to assets (*r*) or by simple transfers (*T*). It may also increase income by reducing the impact of negative shocks (*S*). While increases in assets and returns to assets are pointing to a sustainable growth over time, transfers are clearly not (as they are dependent on continued, external provision). Changes in returns to assets can be detected by estimating shifts in production functions. For example:

$$y = a + rA + r_pPA$$

39. The project is successful in increasing returns to assets if $r_p>0$. This can happen, in particular, via an increase in labour productivity as working efficiency should increase because of an improvement in health and education. Returns like r_p can be estimated within agricultural production functions as well as within health, nutrition or mincerian production functions. In all cases, the project should increase

[15] While most MVPs run for ten years, the specific case of the northern Ghana MVP will run for five years only.

[16] In the year prior to the commissioning of the MV Evaluation, a workshop involving stakeholders, policy makers, DFID staff, and members of the MVP took place in Kumasi, Ghana. This outlined key evaluation questions, and set the framework for the Terms of Reference. Several options were considered for evaluating the impact of the MV, including randomization, a step-wedge design, and a difference-in-difference (DD) approach. DD was deemed to be the best option, as both randomisation and a step-wedge design would require modifying the MV design and implementation used elsewhere in sub-Saharan Africa. Using the DD design allows the long-term effect of an MV to be evaluated with the least changes in its standard design.

returns to factors for a given level of factor employment. Again these changes can be analysed in a DD framework to assess their evolution over time.

40. Estimation of production functions from household survey data is notoriously difficult and we will experiment with several approaches depending on the quality of the data on output and factor prices obtained. One approach introduced by Yotopolous and Lau (1979) consists of the estimation of farm profits as functions of inputs and outputs prices. Supply and demand functions are then derived from the normalised profit function. The availability of full information on farms production and costs however, permits a direct estimation of the systems of output supply and factor demand. Methodologies for system estimation of output supplies as functions of prices and fixed factors are summarised in Sadoulet and de Janvry (1995). Alternatively, Cobb-Douglas production functions can be estimated of gross farm output value on factor use and of gross value of production per hectare on factors and crop dummies.

41. An alternative way to analyse the long-term impact of the intervention on income is by analysing household savings. One condition for long-term economic growth is that households should save resources for productive investments in physical or human capital. The survey tool collects some data on savings but rural households also save by purchasing animals or other valuable assets. One way of looking at household savings consists of simply calculating the difference between income and consumption: *s=y-c*. The project can increase savings in many ways as can be seen by looking at the standard Euler equation of the canonical inter-temporal consumption model:

$$u'(c_t) = \frac{1+r}{1+\rho} E u'(c_{t+1})$$

42. The Euler equation equates the marginal utilities of consumption between future and present period. An increase in savings in the current period will occur with an increase in the returns to savings (r) a reduction in the inter-temporal discount rate (ρ), an increase in expected income via an increase in levels of earnings or via a reduction of their variance. It is often assumed that for poor rural households, ρ is large and that households are de-saving. In this formulation of inter-temporal consumption behaviour, with the exception of the effect of *r*, households are saving for precautionary reasons. The project should reduce precautionary savings by increasing expected income and by reducing its variance. The project should also increase the propensity to save by reducing the time discount rate – the latter effect is a result of the increase in the survival expectations in the family.

43. The patterns of consumption and savings are not easy to interpret however. For example, Chen et al. (2009) found an increase in incomes but not in consumption during five years of implementation of an anti-poverty programme in rural areas of China. Apparently households interpreted the change in income as transient and saved for precautionary reasons. Incomes and savings increased during the implementation phase, but in the period following the completion of the programme only modest income effects were observed while savings decreased. Direct estimation of time preferences and expected incomes may therefore help in interpreting saving dynamics.

44. Perhaps a more efficient way of analysing savings consists of looking directly at the patterns of asset accumulation. As in the income dynamic equation above, assets in the current period are a function of assets in the previous period and of a set of initial conditions:

$$A_{it} = a + bA_{it-1} + \boldsymbol{cZ_0} + dPT$$

45. A poverty trap implies that *b*>1. The coefficient *d* measures dynamics of assets in project areas and tests the occurrence of asset accumulation. This type of analysis can be conducted parametrically or non-parametrically (Barrett, Marenya et al. 2006). Applications to the accumulation of livestock in Madagascar and Kenya have proved particularly successful (Lybbert, Barrett et al. 2004). A semi-parametric estimation of a similar model can be found in McKenzie et al. (2006).

46. In summary, in order to test the project ability to break the poverty trap, we propose to conduct a joint analysis of income, consumption and saving patterns:

- DD analysis of income, consumption and savings over time
- Dynamics of asset accumulation
- DD estimation of returns to factors
- Estimation of time preferences and income expectations

Difficulties in applying the DD approach

47. As discussed in the previous section, the evaluation will use the DD approach by comparing the change in outcomes in MV project areas before implementation to post-implementations, to changes in the same outcomes with an explicit control group. DD allows us to isolate the MV project impact on the outcomes, from the effects of other variables that are changing over time. In regression form the DD impact of the project is given by $y_{ti} = \alpha + \beta P + \gamma T + \delta PT + \eta$ where y_{ti} are the outcome indicators for each household *i*, *T*=0 at the baseline and *T*=1 at the follow-up, *P*=1 if the household is in the project group and *P*=0 if the household is in the control group. The coefficient β is the outcome difference between the project and control group at the baseline. The coefficient γ is the outcome difference over time in the control group. Finally, the coefficient δ is the DD estimator the difference in outcomes that can be attributed to the programme.

48. The validity of the DD approach rests on the assumption that assignment to treatment is random. If this is not the case, as in the MV project, endogeneity is likely and great care has to be taken in the selection of identical controls. In particular, DD assumes that factors determining the outcomes change over time in the same way in the programme and control villages. This problem can be addressed by selecting control villages that are comparable on the *levels* and on the *trends* of variables correlated with the outcomes. To this aim we propose to identify the control communities by matching project communities (using one-to-one matching) to potential control communities using a propensity score obtained from community-level variables in both the levels and the trends (see Section 2.3). Note also that the nature of the MV project rules out the adoption of ex-post 'placebo tests'. These tests assess the impact of the intervention on a variable that is known not to be affected by the programme. If an impact is found, than the validity of the controls used is questioned. The MV project is such that it is impossible to find an outcome that will not be affected by the intervention.

49. Another difficulty of a DD design is **differential attrition**. There is a possibility that households will participate in the interviews in different ways in project and control sites because of their expectations or because of the outcomes of the programme itself. In particular, there is a possibility that more educated individuals will move from the project communities to take more remunerative jobs in urban areas. If this occurs, the project and control samples will become unbalanced over time and straight outcome comparisons will be no longer valid. Tracking people after they have moved or obtaining information in relation to their new location and the reason for moving can be used to attenuate attrition bias via econometric techniques.

50. A further difficulty of DD analysis is **serial correlation**. Serial correlation results from unobserved factors affecting the outcomes that are themselves correlated over time. Serial correlation results in auto-correlated errors and invalid standard errors. Serial correlation does not affect the estimation of the effect size of the intervention, but it affects estimated standard errors and the derived statistical tests. It may affect the rejection or adoption of the null hypothesis in undesired ways. In other words, it may lead to erroneously finding or not finding a statistically significant impact of the intervention. Bertrand et al. (2005) illustrate what econometric techniques can be employed to deal with serial correlation in the context of DD analysis.

2.2 Addressing the core evaluation questions

This section explains how we will address each of the evaluation questions outlined in the ToR. Each question is taken in turn, covering the impact on the MDGs and synergistic effects, externalities and unintended consequences, sustainability, and cost-effectiveness.

Impact on the MDGs and synergistic effects

51. The standard DD analysis will be used to assess the project impact on the MDGs such as poverty, under-nutrition and child mortality. There is however a difference between how this data will be used by the MVP implementation team, and the independent evaluators. The project implementation team will be able to use the baseline and repeat survey data to track progress against the key project, and especially the MDG indicators. This is viewed as mainly a monitoring function that will allow the implementation team to assess the project against indicators for poverty and nutrition, primary education, child health, maternal health, HIV, TB and malaria, and environmental health.[17] The evaluation will also utilise the same dataset, but will use it to undertake a DD analysis to measure progress towards the MDGs, and to measure the net effect that is attributable to the MVP. This analysis will focus not only on a number of key outcome/impact variables, as listed in Table 1, plus also the intermediate outcomes that help inform how these final outcomes/impacts are being achieved. In this way, the analysis will not be limited to the project's own indicators and certainly not to just the targets of the MDGs. This is important as the MDG indicators are not all at the same level in the results chain hierarchy. For example, MDG 1 (to eradicate extreme poverty) relies on an income-based indicator (dollar per day). In order to achieve increases in income there may be any number of *intermediate* outcomes that contribute, from sales of farm produce to off-farm income generation. It is thus necessary for the evaluation to also focus on intermediate variables that precede the achievement of this particular MDG. In contrast, MDG 2 (to ensure that children everywhere are able to complete a full course of primary schooling) has indicators that focus on enrolment. Enrolment and attendance can be seen as occurring much earlier in the results chain, and so the evaluation will also need to focus on variables that are a *consequence* of improvements in enrolment – such as the quality of education and people's learning abilities, using data from cognitive tests.

Table 1: Key outcome variables of the evaluation

1) Poverty	Household income and expenditure, FGT poverty measures, MPI, assets and living conditions
2) Child development	Cognitive tests (Raven's and digit span)
3) Schooling	School attendance, Maths and English test scores
4) Anthropometry	HA, WA and WH Z-scores
5) Malaria	Fever episodes and haemoglobin

52. A key measure for the evaluation is poverty. The poverty analysis will be conducted by employing three different welfare indicators: (i) monetary poverty, (ii) a multidimensional poverty index (MPI), and (iii) asset indices. Poverty indicators will be calculated using household expenditure measured adopting annual recall. The poverty indices employed will be the poverty headcount, the poverty gap and the square poverty gap as in the standard FGT formulation. Poverty lines will be obtained from the Ghana Statistical Service (GSS) in order to compare rates to official statistics.

53. In order to take into account the multidimensionality of poverty, we will calculate an MPI using the dual cut-off method of Alkire and Foster (Alkire and Santos 2010). The data collected in the field allow the computation of an index similar to the Oxford-MPI. The 10 items to include are:

[17] A full list of MVP indicators is for example set out in Table 2 of Pronyk et al. (2012).

- No household member has completed five years of schooling
- Any school-aged child is not attending school in years one to eight
- Any child has died in the family
- Any adult or child malnourished
- The household has no electricity
- No improved toilet or if toilet is shared
- No have access to clean drinking water
- Household has dirt, sand or dung floor
- They cook with wood, charcoal or dung
- Household does not own more than one of: radio, TV, telephone, bike, or motorbike

54. The advantage of the Oxford MPI is that it is a summary measure of achievements of the MDGs. It measures the levels and the incidence of deprivation at the same time, and can be used to analyse the patterns of deprivation and monitoring progress on several dimensions at the same time. Asset indices will be built based on the rich data collected from the household survey on durables and productive assets. The items included in the index, and possibly the weighing scheme adopted, will be selected based on insights from qualitative work. Alternatively, factor analysis will be employed.

55. The real challenge consists of estimating the synergistic effect (something similar to achieving economics of scope) of the intervention, generated by the simultaneous investments in all sectors of the village economy. The project is designed to break the poverty trap, which implies generating outcomes in excess of the outcomes produced by the sum of the single project components. Put differently, the project should be cheaper than other comparable intervention in achieving the same unit outcomes (for example, one life saved). We will address the estimation of the synergistic effects in the cost-effectiveness analysis (see Chapter 3, pages 38-47, for details).

Externalities effects and unintended consequences

56. There are three types of externalities that the MVP intervention may generate. First, there are benefits spreading from the MV villages to the nearby villages. Second, there are likely effects on expenditure decisions and the allocation of public services in the districts; and, only districts where the MV intervention is implemented will be affected by these externalities. Third, there is a possible 'imitation' effect at the institutional level, whereby other districts adopt policy packages similar to those adopted in the MV districts, for example, the free distribution of bednets or fertiliser.

57. The project is likely to have a number of positive externalities to nearby villages. The most obvious effect consists of villagers from nearby localities accessing the services provided by the programme. There are many localities that are within walking distance from the project localities and some 'leakage' of goods and services to these localities is inevitable. Other positive externalities will arise from health interventions such as vaccinations, de-worming and the distribution of bednets. The reduction of infection risk will benefit non-project villages as well. Finally, there are externalities that operate through markets, such as where the programme may have an impact on local labour demand or prices.

58. In order to detect this set of effects, the evaluation has stratified the sample of control communities by distance; thereby building in an additional treatment arm where the control communities are in 'faraway' as well as 'nearby' communities. In each district, half of the control localities were selected from an area within 10-15 km from project localities. The other half was selected from the area beyond the distance of 10-15 km from the project localities. We decided to employ an administrative cut off by defining some area councils, and all villages therein, as 'far' and others as 'nearby'. The assignment of area councils to 'near' and 'far' areas was based on visual inspection of the map and application of the 10-15 km rule above. Geographic distance alone however is not a precise indicator of potential externalities. First, the distance between villages is not a correct representation of the distance between a household and a service point because households are geographically dispersed within

villages. Second, the geographic distance does not reflect the social distance which may exist between localities on the basis of ethnicity or family ties. In order to address the first issue we will employ GPS coordinates at the household and service facility level. To address issues of social distance we have designed a short social network questionnaire that captures the levels of ties between families across villages.

59. A second type of externality originates from changes in the allocation of public and project expenditure by NGOs determined by the programme. It is difficult to predict whether on balance this externality is positive or negative. District assemblies and the government are committed to invest in the MV villages to match the project investment to some extent. If the public budget is fixed this implies a reduction in public expenditure in non-MV localities. On the other hand, many public investments had been planned long ago in the MV villages and therefore the displacement effect is less than the size of the investment requested by the MV project. In addition, other public or private initiatives may in the future prioritise non-MV (as opposed to MV) areas because the area is already privileged by the MV project.

60. In order to detect these effects we have included two additional modules in the community questionnaire. The first module collects data on projects by any funding body in each locality, while the other module collects data on district assembly projects and expenditures in each of the project and control locality. This information will be complemented by qualitative work directed to ascertain how public expenditure decisions at the district level are made, and on what is the likely size of the expenditure displacement. If a rate of return to public investments can be calculated, a correction to the DD overall impact of the intervention can be performed after accounting for public-expenditure related externalities (Chen, Mu et al. 2009). Note however that the population covered by the intervention is 12% of total population in Builsa and 9% of total population in West Mamprusi. This suggests that the displacement of funds by public and private projects produced by the MV programme should not be too large.

61. A third type of externality arises at the institutional level. By demonstration and learning effects, district assemblies throughout the SADA region may put in place policies that proved successful in the MV sites. SADA authorities have expressed interest in learning from the MV experience and in changing policies according to the lessons learned in the field. These effects do have an immediate impact on the evaluation design if policies similar to those implemented in the MV sites are implemented in the selected control areas. More generally, these are positive project effects if effective initiatives are successfully replicated by SADA in other areas. We will document this development by collecting district-level budget data in the intervention and control districts and by conducting a qualitative institutional assessment involving SADA authorities.

Sustainability

62. The evaluation plans to explore the analysis of sustainability in two ways. First, by re-surveying project and control villages after 10 years, i.e. five years after the project intervention has ended. This requires re-interviewing the same households and individuals employing the same questionnaires. There is a risk that the original sample of households will become less representative as, over a period of 10 years, individuals will migrate to other areas or will become uninterested in the survey. It is indeed likely that MVP villages, given the nature of the development process, will produce different migration dynamics with different welfare outcomes and the evaluation team will explore the possibility of tracking migrants outside the villages or refine the existing migration module. If attrition is high and resources for running further household surveys are not available, the evaluation plans to link up with the ISSER-Yale panel survey to provide a good nationally-representative comparator group to track key indicators over time and limiting the survey work to the MV sites only.

63. Second, the evaluation will assess sustainability by capturing changes in irreversible welfare indicators which are known to be good predictors of long-term welfare. In particular, improvements in children's cognitive abilities and the reduction in stunting prevalence will be used with this aim. In addition, the

empirical analysis suggested above (see equation 2 and following discussion) will focus on dynamics of asset accumulation and the changes in returns to physical and human assets, precisely with the aim of separating out the simple impact of transfers and capturing the long term poverty reduction impact of the programme.

Cost effectiveness

64. An important question for MVP is that the synergistic components of the MVP lead to a cost-effective programme. To assess the cost effectiveness of MVP, we will consider the feasibility of a Cost Benefit Analysis of the entire programme, as well as a cost-effectiveness analysis with the use of appropriate non-monetised indicators, and a multi-dimensional poverty measure. As part of this exercise we will conduct a cost-effectiveness comparison of MVP with other relevant programmes; and, have suggested a number of studies such as the Livelihood Empowerment Against Poverty Programme (LEAP) impact evaluation (ISSER and University of North Carolina) and the ISSER/Yale University research study carried out in 2010 in the three northern regions for which datasets are available in order to facilitate this analysis. See Chapter 3, pages 38-47, for full details.

2.3 Key measurement challenges

65. In addition to answering the questions above, there are a number of measurement challenges to be addressed by the evaluation, including the heterogeneity of impact, the scale dependence of programme impact, and the dualism of development process. These are discussed in the section that follows.

Heterogeneity of impact

66. We will investigate the heterogeneity of the programme impact along geographic and household characteristics. The Builsa and West Mamprusi districts are different in many socio-economic respects and are differently administered. A disaggregated analysis of impact in the two districts is required, and the sample of control villages is stratified in such a way to allow this type of analysis. A further disaggregation within districts is unlikely to yield statistically significant results because of the small number of observations. However, two other geographic distributions will be considered across districts. The first distinguishes between remote and centric villages with respect to main markets and service facilities. Project returns to spatial initial conditions will be thus analysed. The second separate a core and a periphery of the project area as by standard coordinates. Because of externalities spatially generated, the geographic centre should benefit more.

67. Household characteristics include gender, but also landholdings and education of the head of household. The impact of the intervention will be disaggregated by subgroups or by quintiles of the relevant characteristic. Non-parametric, semi-parametric and switching regression models will be employed in the analysis.

Scale dependence of programme impact

68. The programme is implemented in a small geographic area and promises to produce dramatic improvements in living conditions. If expectations are met, the programme could be scaled-up to a larger area or region. If the programme is successful, the government could consider scaling it up to the whole of northern Ghana and beyond, in order to put an end to poverty. The evaluation will produce estimates of the impact of the intervention that will help with making such decisions.

69. It should however be noted that standard impact evaluations are not able to provide the information required to make such decisions with confidence. This is a problem of external validity and running a randomised experiment would be no help. The problem arises because of the scale economies generated by the programme. Many of the MV health interventions produce effects that are scale-dependent: vaccinations, bednets etc. The impact of these interventions increases with the number of

people and the size of the area attended. Similarly, the expansion of the agricultural interventions to a large area would produce general equilibrium effects on prices and labour demand that are now localised or diluted. This implies that the impact effect assessed by the intervention cannot be extrapolated to a larger area with confidence. The impact of the MV programme is scale-dependent. As it misses increasing returns, the evaluation will *underestimate* the impact that the programme would have if it were scaled-up.

70. With the exception of theoretical work conducted by Manski (2009), we are not aware of studies that are dealing effectively with this issue. In the course of the study we will investigate methods that could possibly address the problem of extrapolation under scale-dependence. One possibility is to compare a 'core' and a 'periphery' of the MV cluster. If scale economies operate spatially, the different impact in the core and the periphery may provide an idea of how the impact increases with the scale of the programme. Note that scale economies might also be generated in the provision of services. The unit cost of the intervention should decrease with the scale.

Dualism of development process

71. Canonical models of economic development (see, for example, Dixit 1973) predict that in the course of development, countries undergo a process of sectoral transformation. Technological progress in agriculture generates an agricultural surplus; as by Engel's law of demand the increase in the consumption of food is not proportional to the increase in income. As a result, with technological progress, agricultural prices fall and labour demand in agriculture decreases. If working opportunities outside agriculture are available, people migrate to urban areas. The process can be accelerated through heavy investments in human capital as shown for the South of United States by Caselli and Coleman (2001).

72. The effects of the MV project on migration outside agriculture are hard to predict. The introduction of improved technology will increase labour and land productivity and output. The impact on wages and prices is however unlikely to occur given the small size of the intervention. If markets are poorly integrated, local agricultural prices may decrease and wages increase, thus retaining the labour force. On the other hand, the increase in the skills and efficiency of the labour force through investments in human capital may increase the number of people moving out of the unskilled agricultural sector.

73. Migration out of (as well as into) the MV villages may have important implications for the estimation of programme impact. Many of the project benefits may go unobserved if high income earners move to urban areas, or if the MV site attracts in-migration due to the relatively high level of investment. In order to avoid this bias, the in/out migration process needs to be documented, and in particular, earnings of former villagers needs to be tracked. To do this, the evaluation has included extensive sections in the household questionnaire to collect information on household members living elsewhere and remittances. In addition, the evaluation will track household members in their new residence by using mobile phone numbers to the extent this is possible.

2.4 The sampling strategy

74. Based on power calculations, the evaluation will conduct three full-length survey rounds over a five-year period using average samples of 20 households per cluster in the 34 MV localities and in 68 other matched control localities. The study will follow three treatment arms of equal size: MV localities, 'nearby' control localities, and 'faraway' control localities. MV localities are the project group comprising communities where the intervention is implemented. The 'nearby' localities are a control group which could benefit from the intervention because of the geographical vicinity to the project sites. The 'faraway' localities comprise a 'pure' control group that is not expected to benefit from the MV intervention in any way because of its geographical distance from the project sites. The sample size for each treatment arm is 750 households. This is the sample that will be used in the baseline (2012), in

the 2014 round (two years after the intervention), and in the 2016 round (four years after the intervention).

75. It was estimated that a sample size of 20 households per cluster was needed, corresponding to 680 households in the project group and 1,360 in the control group. However, it was decided to 'round-up' the numbers to 750 and 1,500 respectively considering that some communities could be as small as not to have 20 households available for interviews and that some households would be lost over time due to attrition. The sample size for each treatment arm is therefore 750 households. There are therefore 68 control villages and 1,360 control households. This is the sample that will be used in the baseline (2012), in the 2014 round (two years after the intervention), and in the 2016 round (four years after the intervention). See Table 2 below.

76. It is also planned to conduct two additional (smaller) survey rounds on a restricted sample of 15 households per cluster (drawn from the original baseline sample). This will administer the expenditure section of the questionnaire together with some core MDGs outcomes like school attendance during the 'off-MV' survey years. If additional resources become available, the same sample of 750 households per treatment arm will be followed for the entire duration of the study.

77. In the survey rounds of 2012, 2014 and 2016 the full MVP package of questionnaires will be administered by the Earth Institute to track progress on the MDGs. The additional survey rounds of 2013 and 2015 will administer a restricted version of the MVP household questionnaire focusing on tracking progress in poverty reduction. The power analysis shows that income and non-income related outcomes require different sample sizes and repeated measurements. In particular, several MDGs outcomes, like nutritional status of children and learning skills, are highly auto-correlated over time and require larger sample sizes with little benefit obtained from repeated measurements. On the other hand, poorly auto-correlated outcomes, like income and expenditure, greatly benefit from repeated surveys at the same time allowing smaller sample sizes.

Table 2: Sample size of each treatment arm in the six survey rounds (households)

	2012	2013	2014	2015	2016
Treatment sample	750	525	750	525	750
Control sample	1,360	1,020	1,360	1,020	1,360
Total sample	2,110	1,545	2,110	1,545	2,110
Modules	Full MV module	Expenditure module	Full MV module	Expenditure module	Full MV module

Power analysis

78. The sample sizes of Table 2 were obtained from two power calculations. The first calculation is based on the standard DD, consisting of one baseline and two follow-ups. The second calculation is based on a longitudinal study consisting of one baseline and four yearly follow-ups and relies on the analysis of covariance (ANCOVA) by assuming that effect sizes will be calculated controlling for initial values of the outcomes in order to adjust for differences between treatment arms before the intervention.

79. The values used in the power calculations are presented in Table 3. Per capita expenditure mean and standard deviation were calculated from a sample of 600 rural Ghanaian households (GLSS 1998/99) residing in the Northern region of the country. A sample of rural Ghanaian households (GEIES 2003) provided means and standard deviations of height-for-age Z-scores among under-five rural children and scores on a simple maths test among rural children under 18 over a 0-8 point scale (observations from all countries were used in this case because of the small number of observations available from the Northern region). The intra-cluster correlation coefficients were calculated from the same datasets employing the *loneway* command in stata. The autocorrelations coefficients of the selected indicators for Ghana were suggested by the various sources quoted in McKenzie (2012).

80. The evaluation focuses on four outcomes: agricultural income, expenditure, height-for-age Z-scores and maths skills. These four indicators summarise a large portion of the MDG indicators and overall human development. Income and expenditure have low autocorrelation, test scores have an average autocorrelation while the autocorrelation of HAZ (Height-for-age Z-score) scores is generally high. The autocorrelation values are used to calculate power of a standard DD (one baseline and two follow-ups) and an ANCOVA design with one baseline and four follow-ups. The calculations are based on a fixed number of 35 clusters (m) because this is the number of localities that were selected for the intervention. We conduct sensitivity analysis for varying samples of 15, 20, 25 and 30 households per cluster. The calculations were performed using the *sampsi* command in stata after adjusting sample size for the design effect produced by the clustering of the sample. Sample sizes of 35 clusters were multiplied by the number of observations per cluster, and were divided by the design effect defined as: *Deff=1+ (m-1) ICC*. In other words, power calculations are performed on samples of households that were downsized in order to account for intra-cluster correlation.

81. Note that while calculations of income and expenditure per capita are based on number of households, the calculations of HAZ scores and maths tests scores are based on the number of children. There is however, a large correspondence between the two figures. The census of households conducted by MVP (the Earth Institute) in the MV sites shows that one person in five is under the age of five and that there are more children under-five than there are households (with a ratio of 1:3). This implies that even accounting for refusals and missing observations there will be at least one child under-five measured per household. A similar reasoning applies to the educational tests that will be administered to all children aged six to 15. Therefore, the power calculations presented below equally apply to the number of households or of individuals.

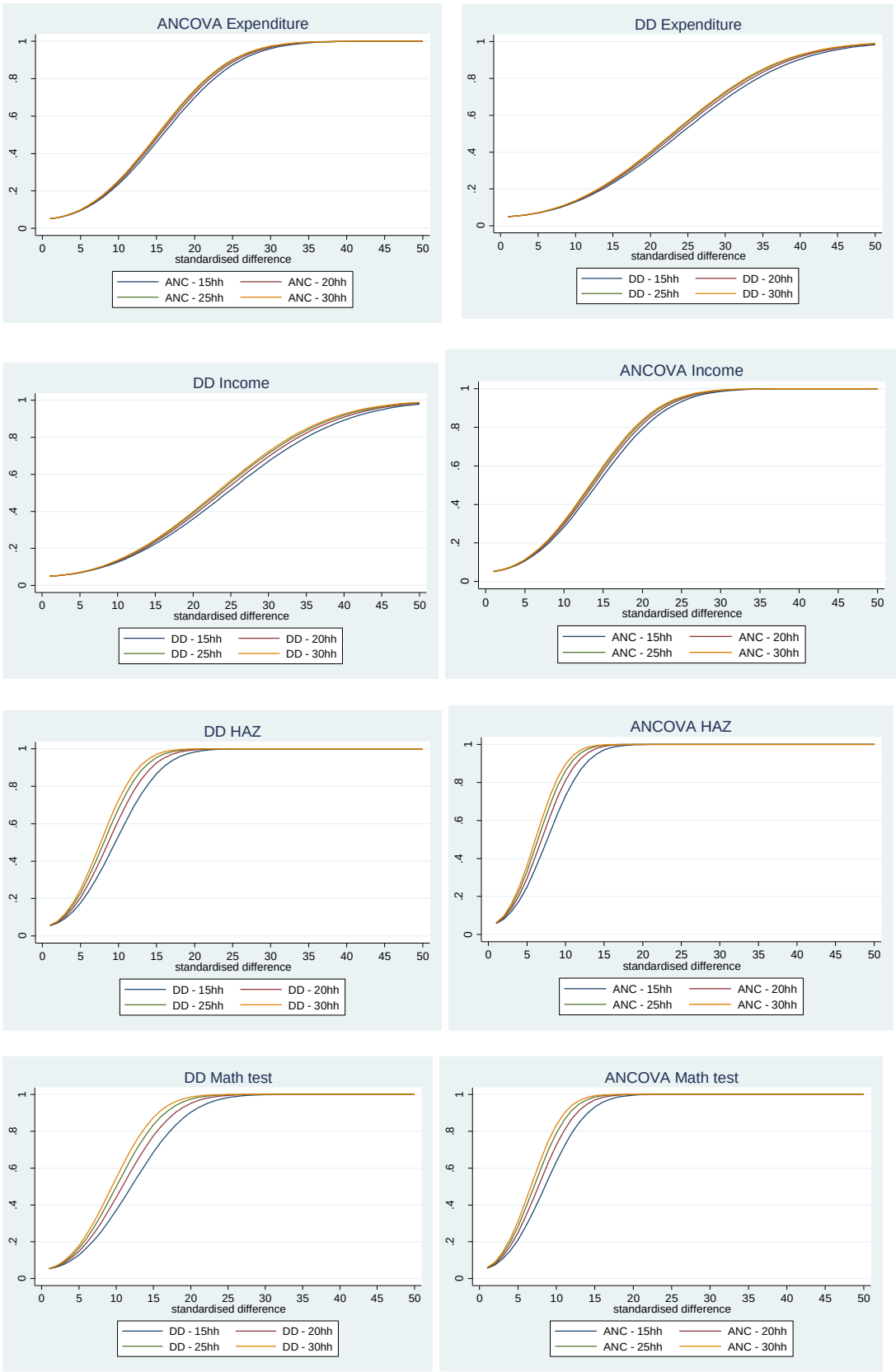
Table 3: Parameters employed in the power calculation

	ICC	Autocorrelation	Ratio mean to s.d.	% equivalent of 0.1 s.d.
Per capita expenditure	0.23	0.40	1.7	6%
Per capita income	0.18	0.25	1.0	10%
HAZ scores	0.03	0.70	0.7	14%
Math test scores	0.02	0.50	2.0	5%

82. Figure 7 plots power against standardised differences for per capita expenditure, household agricultural incomes, height-for-age test scores and maths test scores. The main conclusions of these analyses are:

- For poorly auto-correlated outcomes (income and expenditure) there is little gain in expanding the sample size beyond 15 observations per cluster both in DD and ANCOVA designs
- For highly auto-correlated outcomes (height-for-age Z-scores and test scores) there are gains in expanding cluster size. The gains are decreasing and are rather small for clusters of 25 and 30 households
- Gains to using multiple rounds (only ANCOVA is shown but it would apply to DD as well) are higher for income and expenditure than for highly auto-correlated outcomes. Gains however are large also for HAZ, maths and test scores
- With a power of 0.8, an ANCOVA design employing one baseline and five follow-ups will be able to detect an increase by 20% of agricultural incomes and 10% in per capita expenditure
- A power of 0.8, a simple DD design employing one baseline and two follow-ups will be able to detect an increase by 18% of average HAZ scores and 8% in maths test scores

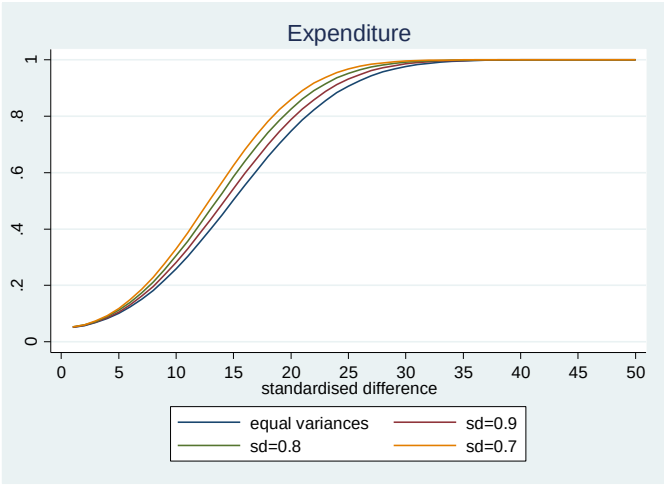
Figure 7: Power against standardised effect sizes under DD and ANCOVA designs



83. The power estimates of ANCOVA were conducted assuming stable variances and autocorrelations across time and across groups. These assumptions might be strong in this specific case. However, a violation of these two assumptions does not play against the selected sample size.

84. The variance of most indicators should be smaller in the MV areas for two reasons. First, the MV sample of villages is likely to be more homogenous that the sample of control villages because of the 'clustering' nature of the MV intervention. Second, the intervention may reduce the variance among indicators by equalising the distribution of outcomes through the services provided. The reduction in the variability of the outcomes however can only have the effect of improving the precisions of the estimates. Figure 8 below shows an example of how the power in estimating differences in per capita expenditure increases for decreasing values of the standard deviation in the treatment group.

Figure 8: Sensitivity of estimated power for decreasing values of the variance in the treatment group



85. The programme could also have an impact on autocorrelations over time. For example, if the programme helps household smoothing consumption, the autocorrelation of per capita expenditure in the treatment group should increase. McKenzie (2012) shows how an increase in the autocorrelation in the treatment group has the effect of altering the optimal distribution of the total sample between the treatment and the control groups. In particular, with higher autocorrelation in the treatment group, a smaller number of observations is required in the treatment group. McKenzie also shows how even large increases in the autocorrelation in the treatment group leads to relatively small changes in the optimal allocation of the sample between groups.

2.5 The selection of the MV (treatment) site and villages[18]

The selection of the MV site

86. The project document prepared by the MPA and the MDG centre, West and Central Africa (WCA) (Millennium Promise Alliance and WCA 2011) lists the following criteria for the selection of the MV site in northern Ghana. The cluster of intervention villages should:

- Be located in the Northern Savannah Ecological Zone
- Cover communities in at least two of the northern regions (Upper East, Upper West, Northern), where incidence of poverty is most prominent
- Cover communities representative of at least one of three broader systems present in the north: (1) river-based; (2) agro-forestry; or (3) growth pole

[18] The districts of West Mamprusi and Builsa have been used in this chapter, as these were the districts at the time of selecting the MV sites and writing this Initial Design Document. Subsequently, the two districts have been split to form the four districts of: West Mamprusi (District Assembly based at Walewale), Mamprugo Moaduri (District Assembly based at Yagaba), Builsa North (District Assembly based at Sandema), and Builsa South (District Assembly based at Fumbisi).

- Include communities that would make a cluster with a maximum population of roughly 30,000 people
- Be a representative site of a typical rural community in the north
- Have strong commitment from local and regional governments

87. Based on the criteria above, three potential sites for the MV project in northern Ghana were initially selected by SADA: The first potential site was located at the intersection of the three northern regions (with no access for four months of the year); the second was located on the intersection of the Northern and Volta regions (a site that had very low population density and plenty of available land that could be irrigated); and, the third site was located in the vicinity of Bolgotanga as a potential supplier of goods and services for neighbouring countries (a potential growth pole). SADA based its initial choice on a 'growth pole' approach. In this approach, areas that show the greatest growth potential are chosen for the intervention. Abundance of cultivable land and water where among the criteria adopted to select the poles.

88. This approach clashes with the 'poverty trap' approach of the MV, which invests in the poorest areas rather than those with greater potential. After consultations between SADA, the Earth Institute and DFID, it was decided to select the first site, but leaving out one of the regions due to population constraints. As such, of the criteria proposed by SADA for selection, only the inter-district nature of the intervention was retained – even though working across two regions is not something that MVP would normally undertake.

89. The project document describes the site in the following way: "*This site includes rural communities located in the Fumbisi valley, Builsa (Upper East) and Kpasenkpe valley, West Mamprusi Districts (Northern Region). It is near one of the major markets (Fumbisi) and is representative of two broader systems: river-based and productive zone. Poverty is endemic, infrastructure development is very limited, and basic social services are lacking. The site has potential for agricultural development as it includes two agricultural valleys*".

90. The choice was guided by criteria of feasibility (for example, working in three districts at the same time as in the Northern-Upper East and Upper West site was ruled out because of the complexity of institutional arrangements) and of development conditions in the area. In particular, poverty status at the district level was a criterion used by SADA. Poverty data at district level are not available, but poverty predictions from the 'small area estimation' based on the census of 2000 and GLSS4 (1998/99) are available. Based on these predictions, several poverty maps at district level have been produced. One of these maps helped the decision on the final selection of the sites.

91. The use of poverty maps as part of the selection process by SADA is problematic. First, the ability of the 'small area estimation' method to produce reliable estimates of poverty rates at the sub-regional level has been questioned (Tarozzi and Deaton 2009). In addition, the poverty predictions for the choice of the MV site were based on data collected more than 10 years before the intervention (the poverty maps were obtained combining data of GLSS4 1998/99 and the 2000 census). However, the biggest problems resulted from more practical issues.

92. The poverty map used by the MV is a map based on 110 district subdivision (Coulombe and Wodon 2007). Though the interpretation of the colour mapping is not easy, this map seems to classify Builsa at a poverty level 10 on a scale from 0 to 10, while West Mamprusi is at level six or seven. A second map by Coulombe and Wodon (2007) based again on a 110 district subdivision presents a different picture. The interpretation of the map colouring is not easy, but Builsa seems now to be ranked six on a poverty scale from 1 to 10, while West Mamprusi is ranked seven or eight. In other words, in the first map Builsa is poorer than West Mamprusi, while in the second map West Mamprusi is poorer. The confusion is not limited to these two districts as all districts appear to have very different poverty rates according to the two maps.

93. The situation is further complicated by the data that generated the maps. Data on estimation of rural and urban poverty at the district level, this time for the more recent 138 district subdivision, produced by the author (H Coulombe 2007), show again a different picture. Rural Builsa has a poverty rate of 57% and is the richest of the 34 districts of the entire north of Ghana (including Northern, Upper East and Upper West regions) with the exception of East and West Gonja. On the other hand, West Mamprusi has a poverty rate of 82%. These rates however are misleading, because according to the Ghanaian census Builsa district is entirely rural as there is no locality in the district inhabited in 2000 by more than 5,000 people (the cut-off point adopted by GSS to classify an area as urban). It is plausible that the low level of predicted poverty in rural Builsa compared to rural poverty in other districts is the result of averaging poverty rates across urban and rural localities.

94. Overall this analysis suggests that: a) the poverty data used to inform the selection of the MV site are dubious; and, b) that these data should not be further used by our evaluation in comparing poverty levels across districts until the major discrepancies between maps and census data are resolved.

The selection of the MV villages

95. Once the geographic area for the MV village was selected (the MV site), all communities within this area were 'potentially' selected for the intervention. Funding is however available for only up to 30,000 households and some inclusion and exclusion choices had to be made. Discussions with project staff suggest that a number of elimination criteria were adopted in the field in order to obtain the final list of MV communities. These criteria were:

- Accessibility: communities far away from main roads and communities likely to be flooded in the rainy season and therefore isolated for a number of months in the year were excluded
- Poverty: communities that after several visits appeared to be considerably richer than others were eliminated. For example, the urban area of Fumbisi was excluded after it became obvious that already enjoyed many of the services provided by the MV project

96. Note that there is a degree of subjectivity in the application of these criteria because they are not based on data on poverty, distance or flooding-risk, but mostly on visual impressions and discussions with local authorities.

2.6 The selection of the control districts and villages[19]

97. The selection strategy of control villages adopted by the MVP in other MV sites can be summarised in the following steps:

- Project staff find areas near the MV cluster that are 'similar' to the MV cluster. Similarity is assessed based on visual inspection and discussions with key informants
- When a similar area is found, a check list of facilities is filled and the area characteristics are compared to the MV cluster characteristics
- The 'potential' control area should comprise six to eight control communities, and three communities out of the six to eight communities are randomly selected to form the control group
- 100 households are interviewed in each control community for a total sample size of 300 households in three communities

98. We consider this strategy as not valid because it does not produce the best possible comparator set of control villages and because of the small sample size. In initial planning for the independent evaluation DFID proposed an alternative strategy consisting of identifying a set of a few clusters of villages and then drawing randomly one or more clusters from this set to build a control group. We opted instead

[19] Please note that the actual names of the selected control villages have been replaced with codes. The names of the councils have been retained. The coding has been inserted into the version of this document that has been made publically available. This is to help protect the robustness of the evaluation, by not disclosing the control sites.

for a strategy consisting of matching project villages to control villages on observable village level characteristics. The potential control villages belong to a large geographic area surrounding the MV site, are dispersed, and do not constitute a cluster as originally suggested by the MVP (the Earth Institute) and DFID. The reasons for this choice are: i) there is considerable risk that a cluster can be affected by an external shock (such as flooding at the time of the survey or a new government programme), while a dispersed set of villages is less likely to be affected by a covariate shock that would affect the validity of the control group; ii) a cluster of villages is likely to be highly homogeneous in characteristics. Statistical power and the possibility of further matching households on characteristics will increase with a control sample with more variability; iii) while the MV cluster can produce economies of scale, there is no reason why economies of scale should take place in a control cluster, hence there is no gain in comparing the MV cluster to a control cluster rather than to a dispersed set of control villages.

99. In addition, the decision was taken to stratify the control villages by distance from the MV cluster, in order to obtain estimates of the MV externalities. Individuals, or even families, residing in villages close to the MV site are likely to move temporarily or permanently in order to access health services and other services offered by the programme. Other general equilibrium effects can be expected on labour supply, wages and prices of agricultural products.

The selection of the control districts

100. The evaluation has considered a number of strategies for the selection of districts from which to draw the control villages, but none were found to be satisfactory. One strategy required drawing the sample of control localities from the whole 'poverty trap' area of northern Ghana including the Northern region, the Upper East and the Upper West. This strategy was found to be too costly either for conducting the surveys or for obtaining data for the selection of localities. A second strategy required matching each of the two MV districts to one or more districts, based on poverty status and other district-level characteristics. The uncertainties related to the existing district level poverty data precluded this strategy.

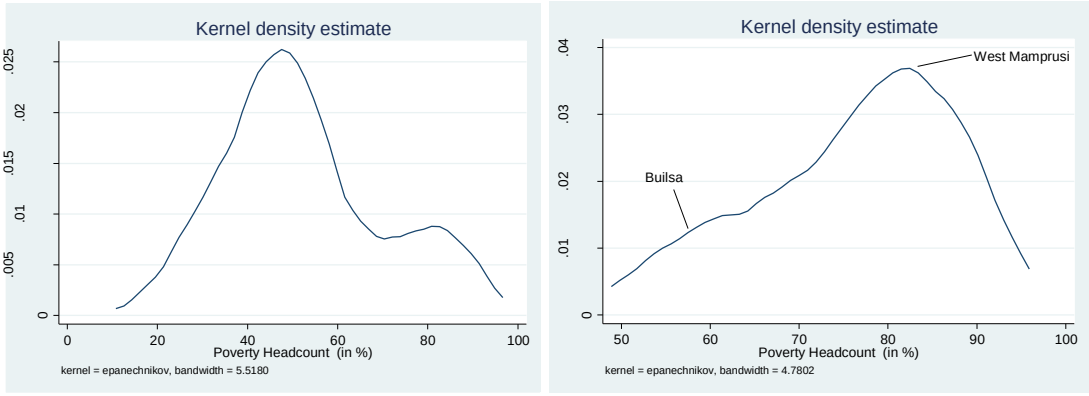
101. The decision was taken to select control villages from the districts of West Mamprusi and Builsa where the project is implemented. The MV area is highly homogeneous in characteristics and selecting control villages from the same area is likely to produce a control group that is more similar to the project group. There were also logistical difficulties in obtaining data from multiple districts and for conducting data collection in a wider area which suggested restricting the geographic area for the selection of the control group.

The selection of control villages

102. The MV project selected 34 localities for the intervention from two adjacent districts (Builsa and West Mamprusi). Unlike other MV clusters, the northern Ghana cluster is not composed of homogenous villages. The cluster was selected in such a way that half of the beneficiary population would be residing in the Builsa district and the other half in the Mamprusi districts. There are no close links or similarities between the villages in the two districts. Indeed the two districts are quite different from each other. Residents speak different languages (Builsa language in Builsa and Mampruli in West Mamprusi). The Builsa are socially organised through clans, while the Mamprusi households respond directly to the village elders. The Builsa area is characterised by good soils and water availability, but soils are poorer in the Mamprusi area and prone to flooding. The Builsa is better served by health facilities with one good health centre in Fumbisi town and several Community-based Health Planning and Services (CHPS) centres in the more remote localities. Settlements in West Mamprusi are fewer and have larger size. Finally, the Builsa district as a whole is richer than the Mamprusi district. Figure 9 shows the distribution of the poverty rates in Ghanaian districts. The chart on the left displays a bimodal distribution that suggests the presence of a poverty trap. Most districts located on the right are from northern Ghana. The chart on the right shows the distribution of the same poverty rates in the

north only. West Mamprusi is in the mode of the distribution while Builsa is in the richer tail of the curve.

Figure 9: Poverty headcounts in Ghana and northern Ghana



103. These differences suggest that suitable control localities should be found separately in the two districts. The project selected 34 localities for the intervention.[20] Localities and population size by district are reported in Table 4. A similar number of beneficiaries is covered in the two districts though the number of villages differs.

Table 4: MV localities and population size

District	Localities	MV Localities	Average population size	Total households
Builsa	183	23	492	11,327
West Mamprusi	154	11	1,134	13,608

Note: figures are those reported by the GSS 2010 census data. The population figures obtained by the census of the MVP are slightly different

104. Each project locality was matched to two control localities within the same district. Matching was performed on a one-to-one basis using the nearest neighbour method. The control villages were stratified by distance to the project site. One of the goals of the evaluation is obtaining evidence of externalities into neighbouring villages. In order to accomplish this task a sufficient number of neighbouring villages needs to be selected and surveyed. In order to ensure this we stratified each district in two areas. The first area comprises the area council where the project is located and the area councils that are approximately within 10 km from the project sites.

105. In the aggregate, if there are externalities, these should be visible in comparisons of the three treatment arms as they should be correlated to physical distance. We acknowledge however that geographic distance calculated in this way is a loose correlate of potential externalities. First, there are practical difficulties in measuring distance of localities many of which are not reported on the map and GPS coordinates are unknown. The 10 km cut off adopted is an approximation. Once data collection is completed GPS coordinates will be known and will be used to verify the validity of the near-far categorisation adopted and any classification error will be taken into account when conducting the analysis of the data. Second, settlements are highly dispersed and the vicinity of two communities (one project and one control) does not imply the vicinity of a household to a project service point. For example, if a project and a control village have each a radius of two km, on average the distance between the centres of the localities is four km but the distance between a household in the control village and the clinic in the project village can be anything between 0 and eight km. Third, social

[20] The actual number of localities selected is 35, but one of these localities (LOC68 in West Mamprusi) consists of only eight households and does not qualify as a census village. We consider this locality as part of the village of LOC69. Hence the total number of project localities from now on is 34.

distance also matters. For example, a control village in West Mamprusi may be very close to a project village in Builsa but more strongly linked to a relatively far project village in West Mamprusi. While the impact of the project will be compared in the aggregate across the three groups, a more detailed analysis will be conducted of the externalities using model employing actual GPS household distances to service points and extent of social links across localities as covariates.

106. Given the circumstances, whereby a randomised trial was not possible and a selection of control villages had to be operated in, in a very short time, we believe the control group built is the most robust possible. 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 difference in difference approach appears to us as the second best feasible approach after a randomised design.

107. The sample of control villages and households obtained is sufficiently large to allow for further propensity score matching at the analysis stage. Once data on villages and households from the selected villages is collected it will be used to drop from the control group those villages and households that are outside the region of common support. Project households and villages will then be matched to corresponding villages and households in the control group using observed characteristics.

108. The matching procedure adopted in the selection of control villages is not perfect. The MV cluster of villages is rather homogeneous and finding good matched control villages is hard. In addition, we had to rely on limited village-level data to perform the matching. Given the time constraints in which we operated, we believe any other solution would have been inferior to the one adopted. Note also that we collected trend data at both household and village level in order to improve the difference in difference analysis and correct for baseline differences in characteristics.

Builsa district

109. The Builsa district was split in two areas. The first area is composed of the following area councils: Chansa, Fumbisi, Kadema, Kanjarga and the Ysobsa electoral area in Wiaga. These areas make the southern part of the district and most of the localities are within a radius of 20 km from any MV locality. The second area is composed of the area councils of Chuchuliga, Sandema, Siniensi and Wiaga (with the exception of Ysobsa). Most localities in this area are far from the MV sites. After consultations with the MV project staff and district assembly, some of the potential control localities were excluded from the selection. In particular, localities of more than 3,000 inhabitants and the communities of EX01, EX02 and EX03, were excluded because they were close to the market centre of Fumbisi or because they are served by good schools and clinics.

Project localities were matched to control localities based on a propensity score calculated from a probit regression. Matching was conducted on a one-to-one basis using the nearest neighbour score for each MV locality. Data for building the propensity score for the Builsa district were obtained from different sources listed in Table 5 but mostly from preliminary census data from 2010 collected by the GSS.

Table 5: Variables used for the calculation of the propensity score in Builsa

Variable	Description	Source
Population in 2010	Number of persons in the locality	GSS census of 2010
Rate of population growth	Annual rate of population growth	GSS censuses of 2000 and 2010
Population density	Ratio of number of households over number of houses in the locality	GSS census 2010
CHPS	Availability of a CHPS health centre in the locality	DHIMS data of the district health service
Primary and Junior Secondary School (JSS) school	Availability of a primary school in the locality	GSS census 2010
Distance to primary and JSS	Distance in km to the nearest JSS school	GSS census 2010
Distance to market	Distance in km to the nearest market	GSS census 2010
Wells	Number of bore and other wells in the locality	GSS census 2010

110. Table 6 shows the output of the probit selection regression. Most characteristics have highly significant coefficients. In Builsa, the R-square of an Ordinary Least Squares (OLS) regression using the same covariates shows that 37% of the variance between localities is explained by the model. MV villages in Builsa have lower rates of population growth (though higher population in 2010), are more likely to have a primary school in the locality but less likely to have a junior higher school and are further away from markets. They also have a smaller number of wells.

Table 6: Propensity score probit regression

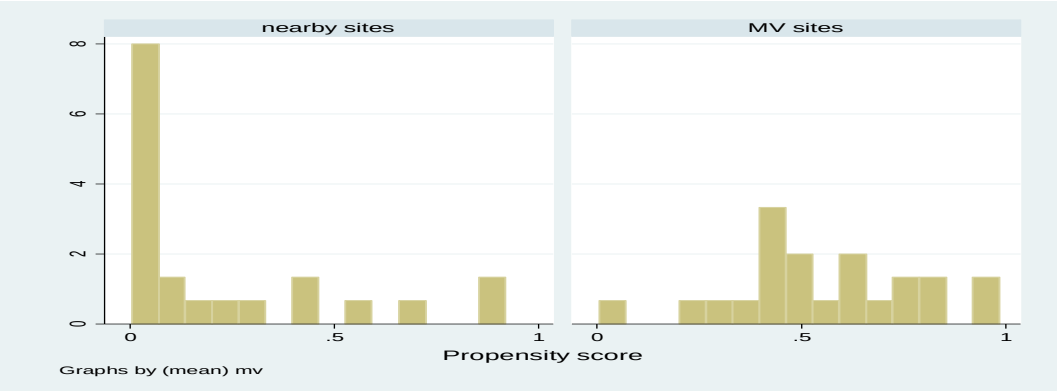
Probit regression — Number of obs = 140
Wald chi2(9) = 37.42
Prob > chi2 = 0.0000
Log pseudolikelihood = -30.926465 — Pseudo R2 = 0.5055

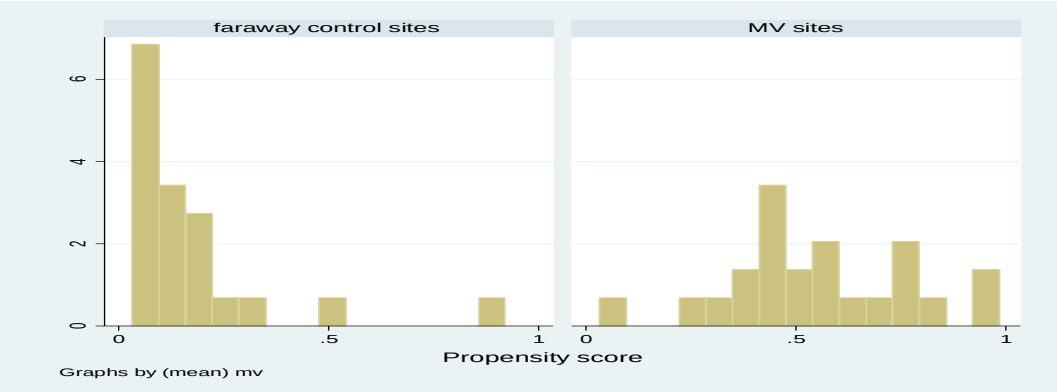
	MV	Coef.	Robust Std. Err.	Z	P>|z|	95% Conf. Interval
Population 2010	.0008027	.0005494	1.46	0.144	-.0002741	.0018794
Population growth	-10.09296	4.154863	-2.43	0.015	-18.23635	-1.949582
Population growth CHPS	.1706194	.0508889	3.35	0.001	.070879	.2703597
Primary School	-2.507032	.6759879	-3.71	0.000	-3.831944	-1.18212
JHS	1.13744	.4719461	2.41	0.016	.2124425	2.0662347
Distance to JHS	-1.228396	.6874809	-1.79	0.074	-2.575834	.1190421
Distance to market	.1742831	.0503961	3.46	0.001	.0755086	.2730576
Number or wells	-.4710694	.2027797	-2.32	0.020	-.8685104	-.0736285
Constant	-4.201246	.7530457	-5.58	0.000	-5.677188	-2.725303

111. The matching of project and control villages was conducted regardless of whether the observations were within the region of common support or not. Several of the control villages and some project villages had very low or very high propensity scores respectively that could not possibly be matched (see Figure 10). However, we decided to retain the villages selected in this way for two reasons. First, similarity in village level variables does not imply similarity of household-level variables, which is what the evaluation is interested in. It is preferable to have a larger sample, though composed of not perfectly matched villages, from which a household level matching can be conducted at the analysis stage. Second, matching of villages was performed adopting the limited number of village-level variables that was available at the time. Once the data collected in the field are processed we will be

able to assess the level of similarity between the three groups considered and we will decide whether any of the project or control villages must be removed from the sample.

Figure 10: Frequency distribution of the propensity scores in project and control villages in Builsa





112. Matching did not remove all the difference between the groups considered but produced a substantial reduction in bias (see Tables C5-C6 in Appendix C). Similarity of characteristics between the project sites and the close-by and far-away sites respectively does not imply a similarity between the far-away and the close-by sites. This similarity is not strictly required because the analysis of externalities will be largely conducted using GPS distance data and social distance assessed by a short social network questionnaire. However, we tested the differences between faraway and close-by villages using a t-test (see Table C9 in the Appendix). Most averages are similar and only one difference (distance to JSS school) appears to be significant at the 10% level. The two samples of control villages appear to be more similar to each other than the sample of project villages to any of the samples of control villages.

West Mamprusi district

113. In West Mamprusi 'close by' villages were selected from within the area councils where the MV villages are located and the neighbouring area councils (Gbmisi/Wulugu, Kpasenkpe, Kunkwa, Wungu, Yagaba and Yzesi). Preliminary data from the GSS census of 2010 were not available with the exception of the population data and we decided to conduct a mini-census of localities in the district after excluding all localities with more than 4,000 inhabitants and localities that were served by a senior high school, hospital or other major health service. This left us with a number of 94 localities that added to the 11 project localities generated a sample of 105 villages. A team of enumerators covered each locality administering a four-page questionnaire to local informants to collect village level characteristics. The exercise, including data processing, took about one week. The variables employed in the selection model are displayed in Table 7.

Table 7: Variables used for the calculation of the propensity score in Builsa

Variable	Description	Source
Population in 2010	Number of persons in the locality	GSS census of 2010
Rate of population growth	Annual rate of population growth	GSS censuses of 2000 and 2010
Population density	Ratio of number of households over number of houses in the locality	GSS census 2010
Builsa	Builsa is main ethnic group	Our mini-census
Mamprusi	Mamprusi is main ethnic group	Our mini-census
Irrigated land	% of irrigated land in the locality	Our mini-census
Wells	Number of bore and other wells in the locality	Our mini-census
Primary and JSS school	Availability of a primary school in the locality	Our mini-census
Distance to primary and JSS	Distance in km to the nearest JSS school	Our mini-census
Market	There is a market within five km from the locality	Our mini-census
Health service	The locality is served by a health service of any type (health centre, clinic or CHPS)	Our mini-census
Floods	Number of floods over the last five years	Our mini-census
Droughts	Number of droughts over the last five years	Our mini-census
Maize	Maize is among four main crops in the locality	Our mini-census
Rice	Rice is among four main crops in the locality	Our mini-census
Millet	Millet is among four main crops in the locality	Our mini-census
Fishing	Fishing is among three main economic activities in the locality	Our mini-census
Trading	Trading is among main economic activities in the locality	Our mini-census
Handicraft	Handicraft is among main economic activities in the locality	Our mini-census

114. The results of the probit selection regression are in Table 8. Many coefficients are highly statistically significant. The R-square of an OLS regression on the same covariate is 0.26. The localities selected in West Mamprusi are slightly less populated and have a larger number of wells. They are less likely to have a Junior High School (JHS) and the nearest JHS tends to be further away. They are more likely to have a market in the vicinity, but are less likely to have a health service of any type. They are more likely to be flooded and historically less likely to be affected by a drought. Interestingly, farmers in the MV localities are more likely to grow millet and less likely to grow rice or maize.

Table 8: Propensity score probit regression (West Mamprusi)

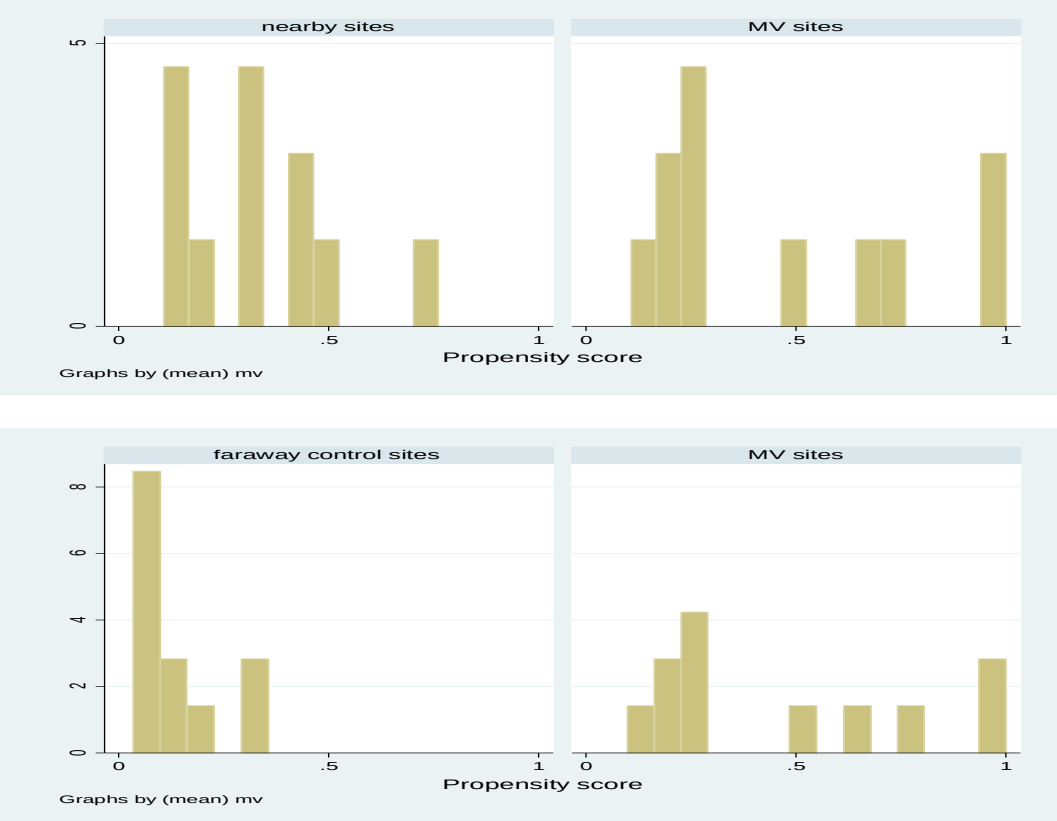
Probit regression — Number of obs = 105
Wald chi2(19) = 30.17
Prob > chi2 = 0.0496
Log pseudolikelihood = -17.995637 — Pseudo R2 = 0.4890

	MV	Coef.	Robust Std. Err.	z	P>|z|	95% Conf. Interval
Population 2010	-.0010198	.0004349	-2.34	0.019	-.0018721	-0001674
Population growth	-0.824371	6.955841	0.01	0.991	-13.71563	13.55076
Builsa majority	-.2161241	1.261763	0.17	0.864	-2.689133	2.256885
Mamprusi majority	-.0400914	.8928802	-0.04	0.964	-1.790104	1.709922
Share of irrigated land	.0012911	.0107714	0.12	0.905	-.0198205	.0224027
Number of wells	.0632905	.0219346	2.89	0.004	.0202996	.1062815
Primary School	-.7181234	.7898251	-0.491	0.363	-2.266152	.8299055
JHS	3.251072	1.195873	2.72	0.007	.9072049	5.59494

	MV	Coef.	Robust Std. Err.	Z	P>|z|	95% Conf. Interval
Distance to JHS	.8645906	.3407863	2.54	0.011	.1966618	1.532519
Health facility	-2.546885	.8181049	-3.11	0.002	-4.150341	-.9434289
Market	2.196804	.6514099	3.37	0.001	.9200637	3.473543
Number of droughts	-.3882498	.2119973	-1.83	0.067	-.8037569	.072574
Maize	-2.41828	.7964922	-3.04	0.002	-3.979376	-.5871837
Rice	-.7209565	.6050367	-1.19	0.233	-1.906807	.4648936
Millet	1.872863	.6952043	2.69	0.007	.5102879	3.235439
Fishing	.4419662	.6631548	0.67	0.505	-.8577933	1.741726
Trading	1.479906	.9372635	1.58	0.114	-.357097	3.316908
Crafts	-.5705799	.6623043	-0.86	0.389	-1.868672	.7275127
Cons	-4.296759	1.431255	-3.00	0.003	-7.101967	1.491551

115. Matching was not able to remove all the difference in characteristics between project and control groups, but the bias was substantially reduced (see Tables C7-C8 in Appendix C). Figure 11 shows the density of the propensity scores for the project and the two control groups compared. As in the case of Builsa, several of the selected control villages and some of the project villages are outside the region of common support. But it was nevertheless decided to include all the selected villages in the survey for the reasons expressed above.

Figure 11: Frequency distribution of propensity scores in project and control villages in West Mamprusi



Validation and de-selection processes

116. A validation exercise was undertaken to identify obvious mistakes that occur naturally in the selection process. This involved a crosscheck with MVP personnel operating on the ground and district officials. For example, in the initial selection of West Mamprusi control sites, two were rejected as they were

also project sites. This is an obvious mistake, as the MVP did not follow the administrative definition of villages, so Kunkwa and Garibinska are just one locality for MVP but administratively they are separate entities and so are reported in the census. Electricity was also accepted as a selection criterion. None of the MV sites has electricity, and so it is reasonable to remove all villages with electricity from the control site sample.

117. Other selection criteria and observations raised by MVP personnel and district officials were rejected, such as vicinity to roads, availability of taxis and so forth. These are not considered transparent selection criteria, and would undermine the robustness of the selection. No visual inspection of the control sites was used to overrule the selection – again because this would provide a potential bias and would introduce new variables that were not used in the matching process.

118. As a result of the validation process, in Builsa one nearby control site was removed from the pool of potential control villages because it belonged to the urban area of the well-developed town of Fumbisi, while no substitutions among faraway control villages were made. In West Mamprusi, some substitutions were made because of erroneous census village names or differences in the agglomeration names. Six villages were removed from the potential pool of control villages because they either had electricity (EX04, EX05, EX06 and EX07) or because they were localised in the immediate vicinity of the major road connecting Walewale to Bolgatanga. The final list of control sites is provided in Appendix C.

Limitations of the selection strategy

119. We acknowledge a number of limitations of this selection process:

- The variables used in the selection are few and are often based on preliminary 2010 census data.
- Matching is entirely based on community level data which may not reflect adequately differences between households. Data on enrolment rates or water access at the household level would improve the matching process.
- There might be unobservable differences between the MV and control villages. However, it should be noted that the selection process of the MV sites was not conducted using clearly defined criteria and a whole geographic area was selected independently of specific characteristics of the villages within. Project staff and the district assembly were involved in the selection process by validating the selection made, which mostly consisted of removing those control villages that would have never been selected for the intervention.

120. Two additional steps will be taken in order to reduce bias at the analysis stage. First, once the baseline data is available we will assess the similarity of the selected villages using a full set of village and household characteristics. Note that outcomes will be compared at the household or individual level. The validity of the control group therefore ultimately rests on the ability of matching project and control households, not villages. What is required at this stage is that no fundamental differences exist at the village level between the project and the control group. We believe we have achieved this goal by matching project and control villages and further validating with field visits.

121. Second, we have collected a number of trend data that we will use to perform further testing and matching at the analysis stage. The community questionnaire contains data for the last five years on: main shocks; asset prices (land and livestock), main agricultural and food prices; wages; fertiliser; and expenditures by the district assembly. The household questionnaire includes trends data on: employment and wages during the three years prior to the interview; monthly profits of micro-enterprises over the last three years; land cultivated over the last three years; quantities produced and prices for the two most important crop produced on the previous three years; number and price of livestock owned over the last three years; and household shocks suffered over the last two years. These trend data will be used to test for the equality in trends between project and control households and for matching on the trends at the analysis stage.

Balance check, timing of data collection and seasonality

122. Data in project and control villages were collected at different times. Interviews in the project sites were conducted over the period April to June, while interviews in the control sites were conducted over the period July-September. This resulted in a time lag of about three months depending on the survey instruments between midpoints of the project and control surveys (see Table 9 below).

Table 9: Time lag between surveys in project and control areas by survey instrument

Household survey	3 ½
Adult survey	3
Anthropometric	3 ½
Blood tests	3 ½

123. Because many of the outcome variables of the project are seasonally affected there is a risk of a bias between baseline project and control data. The bias is mostly introduced by the start of the rainy seasons in June which affects patterns of consumption, production, and infections. In particular, three types of bias are anticipated:

- *Time shift bias:* flow variables may increase or decrease over time and therefore comparisons of data collected at different times will be biased. For example, incomes grow over time and therefore comparing information on income (previous 12 months) at different periods will produce different figures.
- *Recall bias:* people tend to forget past events or extrapolate the present to the past. This generates a bias even if the recall period is the same. For example: 'how many ice-creams in the last 12 months?' will generate different answers in August and in January.
- *Time-sensitive questions bias:* in contexts of high seasonality, short recall questions are biased. For example: how many ice-creams in the last 30 days will produce very different answers in August and January.

124. We conducted an assessment of the potential bias by looking at the recall period adopted in different sections of the questionnaires and we concluded that most problems are likely to arise from time sensitive questions. We note in particular:

- Consumption questions adopting a 30-day recall for non-food high frequency items will be biased (the remaining components of consumption are de-seasonalised as well as income data, provided data were collected as in the questionnaire)
- Food security questions are biased (short recall)
- Mosquito net use in the household questionnaire (short recall)
- Questions on fever (malaria), diarrhoea, child feeding, adult feeding and other health related questions of the adult questionnaire (because they employ a recall of two weeks or one day)

125. The seasonal bias introduced in the data by conducting project and control surveys at different times is problematic for two main reasons. First, if the analysis of the baseline data reveals differences between the project and control data it will be difficult to ascertain to what extent this is the result of having selected poor controls or of the time lag in the data collection in project and control areas. Second, DD estimates will be biased. Even if the follow-up surveys were repeated at the same time of the year every year, no reliable estimates of programme impact could be obtained because the seasonal pattern is not constant and the rainy season could start any month between June and August.

126. In order to address this problem the following strategy will be adopted. First, the evaluation will use secondary data in order to assess the size of the potential seasonal bias for key variables in the region. This analysis will let us understand the potential impact of seasonality on the baseline data collected. Second, using secondary data the evaluation will model key variables using a set of household and

village covariates and seasonal dummies. This will allow us to assess what fraction of the difference observed between project and control areas can be attributed to the seasonal lag and to difference in determining characteristics. Finally, if large differences in project and control areas are found then the impact of seasonality can be estimated.

127. Ignoring issues arising from the seasonal impact of collecting data at different times in project and control villages, there are likely to be differences between project and control data at the baseline because:

- The MV cluster is highly homogeneous and matching on characteristics is difficult
- Matching was performed using village-level data, but even perfect matching at the village level is no guarantee of perfect balancing at the household level
- Matching occurred on observables

128. At the analysis stage we will conduct further matching using the survey data and will drop those households or villages that appear to be off the region of common support. The collection of trend data at household and village level will allow us to match on trends of determining variables.

2.7 Threats to validity: Hawthorne and John Henry effects

129. The study may generate enthusiasm or frustration in the respondents. Respondents may alter their behaviour or misreport facts to drive the results of the study in a particular direction. Some of the localities initially selected for the interventions were subsequently discarded while control communities in the vicinity of the MV sites may be aware of the intervention, which could be the source of frustration. Individuals may also alter their behaviour because they are being studied.

130. These effects are more likely to arise in relation to those outcomes that can be affected by service providers. For example, teachers can spend more time in class in order to improve test scores or nurses can visit more households because they know they are being monitored. It is less likely that the intervention will affect outcomes such as income and expenditure. The ideal solution to this type of problem is a double blind trial design, which in the given circumstances, is not possible. The best that can be done at the design and data collection stage is to avoid the creation of competitive behaviour in either the project or the control group by minimising the impact of survey and qualitative work particularly on local authorities and service providers.

131. At the analysis stage we will investigate the likelihood of these effects through qualitative interviews and observations. We will also analyse data on expectations to detect these types of effects. The survey contains questions on expectations of future income, education and survival. A baseline difference in expectations between project and control groups could be the sign of study effects. In addition, observed future outcome realisations may be predicted by earlier expectations in order to see to what extent expectations have been met or to what extent people have overestimated future outcomes.

132. There is also a risk that households may report incorrect information in the hope of increasing the flow of project benefits or fearing that benefits might be discontinued. Enumerators will be instructed to clearly explain the goals of the study before starting the interview making clear that the study makes the data anonymous and that it will not use the data to inform decisions related to allocation of funds to individuals or communities.

2.8 The quantitative survey instruments

133. This Chapter provides an overview of the survey instruments being used to collect quantitative data for the evaluation. The actual questionnaires are provided in Appendix E part 1. The following questionnaires were based on MVP's pre-existing instruments, with revisions to serve the purposes of

the MV Impact Evaluation. These are to be carried out by MVP (the Earth Institute), which is responsible for the M&E function of the MV project:

- **Household (HH) questionnaire**, including additional modules designed by the impact evaluation team on income, expenditure, in/out migration, and social networks. The other modules are mostly focused on questions regarding MDGs achievements (education, malaria, water, sanitation, time use in the home, etc.).
- **Adult (female and male) questionnaires** – developed from internationally accepted standards for Demographic and Health Surveys (DHS) used to calculate child mortality, etc.
- **Facility questionnaire** – a tool on characteristics, staffing and usage of main health and education facilities (clinics, primary and junior high schools).
- **Village questionnaire** – designed by the impact evaluation team to capture village level data on land area, distance to facilities, economic activities, market prices, shocks and development projects. Prices and shocks are collected in the levels as well as in the trends over the preceding three or five years in order to perform a more rigorous difference in difference analysis of project outcomes.
- **Anthropometry module**: heights and weights of all children under-5 are taken.
- **Blood testing**: haemoglobin of all children under-5 obtained by finger-pricking.

134. The evaluation team and the PRG recommended a number of additional modules that the MVP (Earth Institute) is not able to implement, and will be undertaken by the ISSER, University of Ghana. These modules are:

- **Education and cognitive tests for school age children.** These are designed to observe the MV project's impact on learning outcomes, which are not otherwise captured by the HH questionnaire (i.e. which focus on attendance rates, and highest grade achieved). They include Raven's matrices, backward and forward digit span, short and advanced Maths and English language tests.
- **Expectations questions.** These are designed to test people's expectations of survival, income, and education and educational returns – all of which may be impacted by the MV project.
- **Risk and time preferences.** We have opted for using hypothetical lotteries using the matching task method (rather than real rewards) to test people's choices preference between immediate and (higher) delayed rewards.

Household and Adult questionnaires

135. The Household and Adult (male and female) questionnaires were originally designed by MVP (the Earth Institute), and have been adjusted to reflect requirements of the impact evaluation team, and comments from the PRG.[21] This has required a process of negotiation to improve their quality and appropriateness for the purposes of the impact evaluation.[22] MVP (the Earth Institute) has accepted most of the recommendations made by the evaluation team, including new modules and questions (particularly the income, consumption, social networks, migration, and education modules). The addition of an income module is considered essential for measuring the overall poverty reduction effect of MV (MDG One). The migration module was added to capture the dynamics of migration in/ out (and within) the MV site.

[21] See also: "*Accompanying Note for the Submission of MV survey instruments*", Impact Evaluation team, 16th March 2012.

[22] The evaluation team worked on the following instruments: Dr. Edoardo Masset (assisted by Ricardo Santos and supported by Dr. Patrick Nolen) reviewed and proposed general recommendations to the household questionnaire; Dr. Arnab Acharya reviewed and proposed general recommendations for the revision of the adult male and female questionnaires; Dr. Marzia Fontana reviewed and proposed recommendations for the incorporation of gender elements in the household and adult questionnaires. Dr. Maria Muniz of the Earth Institute also worked closely with the evaluation team and revised the questionnaires.

136. We have been conscious of the length of the survey instrument, and have made recommendations to reduce the overall size of the household and adult questionnaires. MVP has decided to retain almost all the original modules and questions of these questionnaires, as they are standardised across countries for MV sites and are said to be necessary for comparative purposes.

137. The questionnaires designed by MVP are standardised in such a way to be applicable to all African countries. A consequence of this design is that lists of food and non-food items, crops, and units of measurement are not specific to the local context and not even specific to the Ghanaian context. However, after our request, MVP (the Earth Institute) has assured us that the lists and codes of the various questionnaires will be made specific to the local context at the time of training and piloting the instruments.

138. The following MV instruments have not been discussed or revised with the impact evaluation team: (i) Facility questionnaire[23]; (ii) Crop yields; and, (iii) Biomarkers. Of these, we have recommended the use of the Facility questionnaire only, because we doubt the reliability of the yields data and the opportunity for biomarkers data (for the reasons stated below).

Education and cognitive tests

139. We will also use skills and cognitive tests for school age children (five to 15 years old). We believe that this is necessary to observe the project impact on learning outcomes, which will not otherwise be sufficiently captured by the survey instruments (i.e. which focus on attendance rates and highest grade achieved). We will employ the Raven's matrices and the maths and reading tests employed by the ISSER/Yale University socio-economic panel survey – as this is tested in Ghana and might allow us to compare the results. Advanced tests for junior secondary school children were designed by the team based on the current class 1 JHS curriculum. In particular, we will administer:

- Raven's pattern cognitive assessment
- Forward and backward digit span test
- Maths questions simple and advanced
- English reading questions simple and advanced

140. There are 2,100 households in the sample. Based on census data, about 30% of the total population in the district is within the age range six to 17 years. With an average sample size in the districts of 6.8 people per household (DHS 2008), this implies that there should be at least two eligible children in each household thus providing a large sample even considering refusals and the inability to contact all children.

[23] The evaluation team has not been able to discuss and revise the Facility questionnaire as this was conducted earlier as part of the census.

Table 10: Sample sizes for additional modules

Module	Villages	Subjects
Education and cognitive tests (Raven's matrices, Maths and English)	102	Approximately 2,000 children age 5-15
Expectations questions (income, education and survival) Lottery games (risk and time preferences)	20	400 farmers

Expectations, risks and time preferences

141. The PRG have strongly recommended that we test risk and time preferences as the project could have a significant impact on these aspects. There are several reasons as to why the MV project should affect time preferences, including decreasing people's 'impatience' by:

- Increasing the investment and bequest motives for saving (investment motivations may make people less impatient)
- Improving survival expectations (people who live longer are less impatient)
- Increasing income and wealth (poorer people are more impatient because need to satisfy basic needs). However, it will make a difference whether people perceive the increase in income brought about by the MV as temporary or permanent
- Improving education (foresight and planning skills are correlated with education)

142. This suggests that in addition to time preferences we need to also collect data on income and survival expectations. The questionnaire on time preferences (Appendix E part 1) also includes questions on income expectations and survival expectations. There are two reasons for this choice. First, by asking the expected agricultural output we obtain a starting amount to be used in the game that is sufficiently large and meaningful to the respondent. Second, we want to be able to disentangle the effect of survival and income expectations from stated time preferences. Every delay implies a degree of risk, which can be related to the probability of occurrence of the outcome or the possibility of enjoying the outcome. In principle this risk component of the choice should be separated out from the elicited preferences in order to estimate 'impatience'. Data on expected incomes are elicited from farmers by extracting subjective probabilities of future events (Attanasio and Augsburg 2011).

143. We have opted for using hypothetical lotteries rather than real rewards because hypothetical rewards have the advantage of allowing the interviewer to play with several amounts, large amounts and different time horizons at the same time. Hypothetical lotteries have the disadvantage of not providing incentives for the respondent to focus on the game, but reviews comparing the results of hypothetical and real lotteries have not found significant differences (Frederick, Loewenstein et al. 2002).

144. We opted for employing the *matching task* method rather than the more common *choice task* method. While in the choice task method respondents are presented with alternative choices, in the matching task method the respondent fills the blank to equate two inter-temporal choices. For example, s/he will state the amount of money in three months that is equivalent to a given amount of money now. This method has a number of advantages. First, with just one answer the indifference point is identified (rather employing multiple questions as in the choice task). Second, there is no anchoring problem because it is the respondent that provides the initial amount. Anchoring occurs when a choice is affected by the previous choice, which is a common problem of choice task experiments (Frederick, Loewenstein et al. 2002).

145. There is a risk in employing the matching task method of obtaining 'coarse' answers whereby the respondent responds quickly applying simple heuristic rules to the sum initially offered. We will try to avoid this effect by designing a game that simulates a real life situation. We will restrict the game to the case of a transaction involving some agricultural output. The amount involved in the transaction is the

expected amount as stated by the respondent. This will make the game realistic and will also remove the 'magnitude effect' by basing the game on an amount that is meaningful to the respondent (the magnitude effect is the bias produced by the fact that people tend to apply larger discount rates to smaller amounts – the reasons for this behaviour are not well understood). The amount initially stated will be probed by bargaining. The interviewer will encourage the respondent to accept a smaller amount until an agreement between the two is reached. This responds to the need of approximating real life price negotiations and to the need of removing unrealistic initial responses. The respondent will be forced to find the minimum amount he is willing to accept in exchange for a delayed payment.

146. The game will be performed over three different time horizons: one month, three months and six months. The reason for repeating the game under different horizons is detecting hyperbolic discounting, whereby people tend to discount more heavily choices that are made over time horizons that are closer in time to the time of the interview. Heavy hyperbolic discounting is a sign of impatience and of poor saving planning skills (Ashraf, Karlan et al. 2006). In order to avoid that the respondent applies simple heuristics to the different choices made (for example, preferring 110 to 100 over a month and 130 to 100 over three months) we will use a titration procedure, whereby the different time horizon will be presented in random order rather than by an increasing or decreasing sequence.

We piloted the experiment in the field and given the simplicity and speed of the exercise we decided to administer to one farmer (preferably the head of household) in each of the 2,100 households interviewed.

3 Cost-effectiveness methodology

147. The cost effectiveness analysis of MVP aims to inform policymakers about whether (or not) it will be fiscally desirable to replicate the programme.[24] As the impact of MVP will be obtained at a considerable cost, the issue as to how the programme compares to similar interventions in terms of cost effectiveness is highly relevant. The programme theory suggests that the MVP should be more cost-effective than other development programmes because of the synergistic impact on poverty alleviation. This stems from the fact that one of the intentions of MVP is to have an economy-wide impact and to affect wellbeing in a multi-dimensional way – whereas most other development programmes often have singular or more targeted aims. Within MVP, for example, farm productivity will increase directly because of technological improvements but also indirectly because farmers are in better health. As such, traditional methods employed in CEA will not readily lend themselves to assessing the value for money (VfM) aspects of the MVP.

148. In this chapter, we present a number of methods for assessing cost-effectiveness: (i) Cost-Benefit Analysis (CBA) of the entire programme; (ii) Cost-effectiveness of each component of the programme; (iii) Cost-effectiveness using multi-dimensional poverty measures such as that from the Oxford Poverty and Human Development Index, OPHI; and, (iv) Simulation for the scaled-up economy-wide impact along with some attempts at local area economy-wide impact that may include the control area. The chapter then goes on to explain the key variables of interest, the requirements for cost data, and potential comparator programmes. The chapter ends with some of the risks associated with this work.

3.1 Approach and methods

149. There are four main methods outlined in the section that follows, and each is discussed in turn.

Cost-Benefit Analysis of the entire programme

150. One of the aims of the MVP is to generate more income for participants, and this of course is directly monetised. The internal rate of return (IRR) calculations from project costs is a common way of assessing projects designed to generate income. Although the translation of health and educational achievements into monetary units can be regarded as unpalatable or unconvincing, it may be that in order to take into account the multi-sectoral aspects of the MVP, a common unit through which all outcomes can be assessed is required. If it is possible to monetise all items in order to construct a CBA *and* find an extremely high internal rate of return,[25] then no costing comparison will be needed as the programme will be considered relatively inexpensive. Achieving an IRR that will be helpful in arriving at such unequivocal judgments is however likely to be slim. CBA can also be used in conjunction with a comparator; the 'big push' to lift people out of poverty should yield a higher return than any other comparator programme.

151. As a heuristic device we will ascertain the extent to which the current state of wellbeing indicators can be monetised during the baseline period. This will show the feasibility of what can be done at the end-line. The effort to monetise the current situation may be more difficult to obtain for a static situation, as most studies can only assign monetary value to incremental changes. For example, an incremental school year can be associated with increasing wage. Further, it is easier to ascertain changes in wellbeing than measuring overall wellbeing. For example, we may not know the current life-expectancy; however, a drop in under-five mortality (U5MO) may give us an idea of how life expectancy will change.

[24] The objective of the cost effectiveness assessment of the MVP is to address the key evaluation question: "*Is the MV intervention package cost effective in the results it achieves, compared with possible alternatives?*" (ToR, Section 4.1 (iii)).

[25] We will leave it to policymakers and our advisors to determine what would count as a high enough return.

152. Cost-benefit is clearly more suited for the components of the MVP that aim to generate income such as agriculture; where the stream of income generated through sales is balanced against expenditure to give an IRR. To those elements which generate income in more standardised ways we must add all monetised value of the programme outcomes that may be considered extra 'welfarist' measures. We also note that the current project may generate future income and benefits; given the uncertainties and the discount rate we may restrict our analysis to benefits accruing in the next 10 years if we find the trends as they are in the first five years.

Cost-effectiveness of each programme component

153. We however expect that a CEA will need to be carried out for each major programme component, as this is likely to yield more useful information for policymakers – particularly, as any attempt at CBA of the entire programme may depend on controversial assumptions. Cost-effectiveness will be estimated by calculating cost-effectiveness ratios: the ratio between the difference in programme costs in the MV area (C_{MV}) and costs in the control area (C_C), and the difference in outcomes between MVP and control areas (the difference obtained by the DD estimator). The CE Ratio or the incremental cost effective ratio is the following:

$$\frac{Costs\ in\ MVP\ Area - Costs\ in\ Comparator\ Area}{Achivements\ in\ MVP\ Area - Achievements\ in\ Comparator\ Area} = CE\ Ratio$$

154. The synergistic impact of the MVP should mean that each component of the MVP will be more efficient than similar programmes that are carried out separately; this will be the distinctive multiplier impact of MVP. Cost analyses are crucial to making any claims regarding synergistic impact of the MVP. Once the MV implementation commences,[26] we will examine whether each major programmatic element has a smaller incremental unit cost than would be yielded by a non-synergistic programme. For example, the costs of retaining an extra student in primary school should be smaller in MV than in control areas. This is because extra earnings from income programmes may result in fewer children working. Thus, more children are enrolled without a stronger campaign aimed at increasing enrolment.

155. We stress *incremental CEA*. This approach views a particular programme in comparison to others as indicated by the CE Ratio. Thus, only the DD evaluation measures of the programme's effect enters into CE Ratio as the benefit. The denominator represents what can be attributed to the projects; it is the impact factor in most cases of singular programmes. Supposing that the adjusted difference between the MVP and the control sites are 10% in school enrolment for one year - a rise of 200 in total numbers and a cost difference of $2000, then $10 is achieved in raising school enrolment for one year. This CE Ratio can be compared to an outside programme in text books and uniforms where $20 may have achieved a gain of one year of school enrolment; these comparative figures would indicate a synergistic impact for MVP. Although a lower CE Ratio could be due to a variety of factors such as better implementation or programme design, we will expect that synergistic effect to imply that for the same amount of money the achievements of the MVP are larger.

156. While an important factor in the MVP is the synergy through which multiple goals can be achieved, the units of measurement are likely to be very different for each component of the programme. For example: (1) Educational achievement will be measured through increased enrolment and fewer dropouts; both can result in higher per capita number of years schooling by a certain age which in turn can be assigned a monetary value; and, (2) Reduction in malaria incidence may translate into fewer number of U5MO.

157. Many of the outcomes of the MVP will need to be assessed in **non-monetary terms**. For health or education programmes we are likely to find acceptable indicators for programme outcomes. Such

[26] Note: We have examined the proposed MV documentation on capturing cost data, but as yet such systems have not been tailored and implemented for the northern Ghana site. For this reason, the cost-effectiveness lead expert completed a field trip to undertake a more complete assessment of cost data in relation to activities and outputs.

measures like poverty reduction rates can also be used as a specific indicator and will be a good non-monetised indication of income wellbeing. Key outcome variables may include: average rise in children's test scores, infant mortality prevented, or drop in diarrheal rate. It is difficult to monetise these benefits. The cost-effectiveness ratio (CE Ratio) will generate a measure average cost for a unit of gains attributable to the intervention/MVP activity.

158. One further issue that will need to be addressed in the CEA is **timing**. In the MVP, costs and benefits will be encountered at different times. All costs and benefits will be interpreted for the policy makers to be decisive about the inception of the project. Thus costs and benefits will be discounted or inflation adjusted to 2012, where budget is allocated for the project.

159. As stated previously, each component of the MVP should be more efficient than what we find in Ghana when programme activities target only one type of outcome. For instance, children should obtain higher marks at lower programme costs than other programmes with the same aim; or similarly, malaria reduction should cost less than other malaria programmes. Each outcome should be associated with a distinct set of activities with a monetary amount to carry out those activities. This may not easily be accomplished when the MVP has multiple goals. Agricultural programmes may enhance health; similarly, health programmes may enhance agricultural production. Costs may need to be divided to reflect the multiple targets for commensurable comparisons with other programmes. An approach suggested by Dhaliawal et al. (2011) is to simply divide all costs by the number of different goals for which the total costs were incurred.[27] In this way, if such a blunt instrument yields acceptable cost-effectiveness ratios for many of the activities then it is possible simply to accept that the programme is cost-effective. Sensitivity analysis can also be constructed around the apportioning to see how well each component stands up to further scrutiny.

Multi-dimensional poverty measures

160. Although it may be difficult to ascertain a comprehensive state of wellbeing from the baseline for the control and MVP sites, we will still be able to ascertain important indicators such as average education and poverty. These multiple indicators can be grouped into a single measure, such as the OPHI which includes a strong gender component. For more details see Section 2.2, page 15. Although multi-dimensional measures can be arbitrary, they do provide a comprehensive assessment that is an alternative to a monetised value of development. Plus, the assumptions needed for CBA may be greater than those required for multi-dimensional indicators, although a CBA requires less value judgment.

161. Suppose a new CE Ratio is constructed using OPHI as the performance indicator. A low ratio is likely to indicate significant success of MVP; further, this will capture the synergistic effect.

Simulation for economy-wide impact[28]

162. As noted in various parts of our evaluation design, we suspect that MVP being only implemented in a small part of an economy would induce impacts such as migration and spill-over effects. When scaled up, the impact may be different. The simulation of impact may help us to delineate what the impact would be of the absence of migration and spill-over effects. A simulation will however go only part of the way to explaining how the MVP impacts on the economy when implemented on a large scale. MVP will impact the economy through improvements in poverty reduction, in health and in the direct formation of human capital. In this way it is possible to assess the impact of achieving the MDGs on the economy. We suggest the use of a tool for economy-wide strategy analysis, such as MAMS (Maquette for MDG Simulations) from the World Bank.

---

[27] Dhaliwal et al. (2011), *op cit*.

[28] Note: The requirement to undertake a simulation of the economy-wide impact was suggested by the PRG. Implementation of the methodology is dependent on additional resources being approved.

163. The World Bank has constructed an economy-wide model using MAMS for Ghana in order to address questions about the effects over time of different government and foreign aid policies on social and economic performance. The MAMS model for Ghana has a social accounting matrix (SAM) which would make the current task significantly easier. MAMS significantly extends the standard computable general equilibrium model to cover the processes that determine the MDGs and educational outcomes and their feedback to the rest of the economy. MAMS views the MDGs as productive forces in the economy; expenditure for the MDGs of course affects government budgets, the balance of payments and GDP through a simple multiplier effect as well as through productivity.

164. Using the DD parameters as effectiveness values and extending the MV data on costs as well as to estimate the costs needed to fill the current gaps in education and health, we can provide the essential parameters for MAMS simulation to estimate economy-wide effects over the near future (as MAMS is designed to offer a picture of the economy over 5-45 years). As the model is dynamic, we will be able to simulate the impact of a fiscally responsible expenditure path using the parameters from the cost and effectiveness data from the MV study. The expenditure path will fund projects that will gradually erode poverty and gaps in health and education levels. Sensitivity analyses would offer policy choices through use of MV projects.

165. Local economy-wide impact evaluation (LEWIE) has been used as a simulation tool where the recipients in a region are not comprehensive; that is, not everyone in the region received direct benefit from the programme such as a targeted cash transfer programme (see Taylor 2012). One can also measure the spill-over impact through this method; as MVP would yield a spill-over impact in the control area, and LEWIE may be used to measure the total impact of MVP. Such simulations require building local regional SAMS for both the MVP and the control areas. The implications for building local area SAMS for costs are high. Once the programme is implemented economy-wide, there would be no spillover impact. In our view, it would be more informative for us to construct the economy-wide impact at a larger scale using the Ghana SAM, rather than building the regional SAM. We may however find that SAMS for the Northern Region actually exist; in that case we can attempt LEWIE.

3.2 Methodological steps

166. Conducting a CEA requires a multi-step process. After defining the outcomes of interest, we will conduct cost assessments of activities that correspond to bringing about each outcome. The CE Ratio that we propose will be calculated by using similar values from programmes aiming to achieve similar outcomes. Further investigation on the feasibility of measuring the CBA for the entire programme, as well as undertaking the simulation method, will be undertaken during the baseline period.

Key steps in undertaking a cost-effectiveness analysis

167. Ideally an evaluation should be able to capture all impacts, intended or unintended. This however is not always possible. During the baseline period, the evaluation team will **identify outcomes of interest** and ascertain what can be expected to be measured. Identification of impact will follow the results of the impact evaluation. The start of each activity must be noted and assigned to expected types of outcomes. An exhaustive account of activities and the associated outcomes is required from the start.

168. Given the theory of change, a set of measures will be considered as the primary endpoints and will be obtained from the surveys (and their definitions of the outcome indicators).[29] The changes in primary endpoints from MVP, will take into account any unintended consequences, will be compared against those found in the comparator sites. The intensity of programme efforts leading to these outcome measures should be noted every year as part of the theory based evaluation. We also will record a response to the programmes as to how well they are received. This may be vital in programme

[29] Note that there is a difference between what outcome variables will be observed every year (income and expenditure) and every two years (education and health).

attribution and our ability to associate costs with outcomes. Thus, aside from data collection on outcomes, a major step will be how well the programme is implemented and puts in place measures that ensure we have data on how well various programmes achieve their intermediary goals. Evaluation activities will also need to include an account of how well different components of the project are implemented and received. Qualitative data will ascertain how well the programmes are used by the community as well as the community's attitudes toward these programmes. We will further ascertain through programme implementation data if the programmes are put into place to the intended extent, how they have been altered and how intensely implemented.

169. **Extrapolating from effect size:** Cost-effectiveness analysis is useful if outcomes can be expressed in terms of strong welfare indicators such as 'death prevented' or 'number of school years gained'. The MVP may not be powered enough for us to observe some important measures such as the U5MO rate, as the MVP is not designed around this outcome. A decrease in malaria incidence or diarrheal episodes however will likely lead to U5MO. Similarly, we may find that there is noticeable improvement in test scores, which increases the number of years of schooling. Outcomes achieved in the MV project will have long-term effects, for example, a pupil's enhanced educational achievement is likely to create greater income in later years. We will use methods to calculate such returns.

170. Existing studies will be used to make extrapolations of effect size that can be measured against outcomes that may not be readily available from our survey but may be of interest to policymakers. For example, Mincerian returns to education may have been calculated for Ghana. An extrapolation that may be difficult to make is determining the general equilibrium effects when the MVP is scaled up. The effect size we will obtain is conditional on whether only a small number of villages are included in the MVP. If the programme were implemented in Ghana to provide benefits for all Ghanaians who are eligible, the impact may be very different. For example, a year of education would yield higher benefits in a country that has lower average level of education; however, what would happen when all Ghanaians have secondary education cannot be predicted from the current data. No such extrapolations will be made.

171. **Aggregation:** The MVP has multiple components which will engender different outcomes with different units of measurement. Although contested and methods can vary to yield different results, it is possible to take measures and derive a monetised measure (e.g. by measuring the increase in years of schooling). Value of life measure is a monetary value that can be used to scale Disability-adjusted Life Years (DALYs) or Quality-adjusted Life Years (QALYs) into monetary units. We will aim to aggregate all outcomes into a single monetised value. As suggested previously, one possibility for achieving this is to create a multi-dimensional measure like the OPHI.

172. The impact evaluation clearly delineates the importance of control sites (as set out in the previous chapter). We will use DD to measure the impact for a variety of key indicators listed below. As we acknowledge in the above section, we need to link specific components of MVP to observable outcomes that can be detected from the surveys; this will be done during the baseline period. Key variables we need to observe can be divided into three categories:

Income and Poverty

1. Per capita expenditure and income, underlying growth rate
2. Total reduction in the number of people living below $1.25/day
3. Productivity increase in key agricultural output
4. Savings rate
5. Hourly wage
6. Changes in the distribution of male/female work hours

Health

1. Diarrheal and malarial episodes for children under five
2. Immunisation rate
3. Anthropometric Measure: this may be one measure of health but does not include those who died
4. Combined DALY measure: anthropometric measures may not translate to a DALY
5. Combined programme prediction of increase in life expectancy

Education

1. Enrolment rate
2. Number of days absent
3. Improvement in test scores
4. Extrapolated wage

173. The first three education variables will be affected by health programmes. We should expect synergistic impact on all these variables. In terms of the poverty reduction variables for dollars per day, these should not be simply viewed as the number of people moving above $1.25/day. Measures such as the Sen Index can measure deprivation as a gap from poverty line. We therefore will attempt to compare the sites at the lower tails of income distribution, and in doing so, we should observe that income growth in poorer groups is higher in the MVP site than any other group; the increase should also be higher in absolute amount not just in percentage terms. This measure can be adjusted to obtain aggregated measures comparing how well the poorest people do in MVP area (see Ravallion 2004).[30]

Data collection on costs

174. The two main activities regarding data collection specific to costing will be:

- Regularly accounting the costs of a singular module within the MVP including payment transfers
- Costs of comparable programmes in control sites by using government data or cooperating with organisations carrying out similar projects in the MVP implementation area. Note: we will use cost figures from other studies (Ghana or elsewhere) when comparing programmes.

175. The first step in costing involves accounting resources. Costing follows after all resources used in the project have been accounted for even if not all resources will be directly paid from the programme budget; that is, all ingredients going into the programmes will be priced. For example, all labour data should be expressed in terms of time usage to ensure that we not only know the price of labour, but also how much was required. We will rely on pricing both at shadow costs and actual costs paid as the actual costs may at times capture costs of current inefficiencies such as incentive compatible costs, which may not be easily corrected in the future.

176. The process needed for capturing the ingredients needed to run the programme will require:

- Accounting all components of the programme
- Cataloguing inputs used for the programme in a manageable way by categorising inputs into types of resource and linking them to outputs

177. In order to calculate ingredient costs we suggest that the recording of expenditure is itemised as much as possible through the delineation of different components of the MVP. It will be impossible to assign shadow costs to the programme if resource use is not well documented. Any items not paid for must be noted for how it was acquired.

[30] Ravallion M, 2004. Pro-Poor Growth: A Primer, World Bank.

178. Protocols for costing health projects are well established,[31] and a similar methodology will be sought for other project components.[32] Some standard approaches to addressing issues relating to costing different inputs are explored below:

- **Fixed costs:** Annuitisation of fixed costs is common. Local measures on depreciation rates and longevity will be used for various capital inputs. A discussion on the discount rate used will be conducted with the World Bank or Ghanaian Ministry of Finance. Uniformity with local standards or other comparative programme implementers, for example the World Bank, will be helpful. In many CBA analyses for developing countries it is common to apply a discount rate of 8-15%. Such high rates for future health outcomes may be ethically unacceptable; thus following World Health Organisation (WHO) guidelines we may discount future health outcomes at three or 0%. An adjustment to such factors as rates of depreciation and discount can be made at the analytical stage, prioritising recording ongoing expenses should be crucial. Thus purchase price and initial outlays for all fixed costs should be clearly recorded.

- **Multiple usages of inputs:** Where possible, labour working on multiple tasks will be asked to fill out time sheets after the start of the programme. All usage of even a single item must be documented. From the onset, we will list all the inputs a particular item has, and attempt to distinguish proportions of usage. For the final analysis we will try a sensitivity analysis around these proportions.

- **Baseline inputs:** There may be many inputs already in place and used on the project. The present infrastructure enabling the implementation of the MVP must be taken into account if there is little use for such infrastructures other than for the implementation of MVP-type programmes. It may be that such infrastructure would be built regardless of the MVP implementation. In that case, these basic infrastructure elements are immaterial toward costing.

- **Fixing currency and time:** We will report all costs in both dollars and Ghanaian Cedi (GHC). Conversion rates will be based on one year. All costs need to be reported at the common year price (index year) to capture inflation. The primary interest may be to report first in GHC. Exchange rate conversions should occur at the current year and inflation adjusted to the index year. As the inflation rate on dollars may marginally affect the programme, we suggest that final costs expressed in GHC will be converted through the exchange rate in the index year.

- **Transfers:** If the MVP makes any conditional transfer payment we should consider that as a cost; this is in contradiction to many programmes where transfers are not seen as a cost to the economy. In the MVP, transfers are an alternative to another feasible project, thus we should include transfers as costs.

The selection of comparators

179. We will assess costs of similar activities carried out in the control area. It is likely that there are regular activities aiming to improve all outcomes of the MVP targets. First, there will be government projects in nearly all control sites for which costs data would be collected by the government. Cost data from the government tends to only report costs to the Treasury, which often underestimates actual costs. Furthermore, it may be difficult to know what exactly has been costed. Thus there is a risk of comparing incongruent cost data. Government accounting processes will need to be clearly understood. If the methodology differs from shadow pricing methods, we will attempt to transform the

[31] Tan-Torres Edejer T, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans DB and Murray CJL, (2003). Who Cost-effectiveness Analysis, Geneva: WHO.

[32] See: Jimenez E and Patrinos HA, (2008). *Can Cost-Benefit Analysis Guide Education Policy in Developing Countries?* World Bank, Policy Research Working Paper No. 4568; Dhaliwal I, Duflo E, Glennerster R, Tulloch C, (2011). *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education*, JPAL, Working Paper.

numbers to reflect shadow prices by reviewing the accounting processes as much as possible. There is some reason to believe that a substantial proportion of government outlays are made for salaries for most programmes. There will also be some capital costs undertaken and recurrent costs to maintain items such as vehicles and offices. We hope there may be a methodological template within the World Bank to assess retrospective costing to reflect actual opportunity costs. If no such material is available we will estimate value of capital outlay using commercial pricing guidelines. Where this is not possible we will carry out the sensitivity analysis.

180. The second source for the comparator can be similar programmes near the MVP sites conducted as experiments or pilot programmes. These programmes may know the costs of all programmes they are running, including the baseline programmes in place, which may be in addition to programmes already running. For example, potential comparable programmes are conducted by the Jamal Poverty Action Lab (J-PAL) through Innovations for Poverty Action (IPA). The evaluation may be able to use the results of programmes aimed at reducing ultra poverty as well as those increasing agriculture production and improving educational and health achievements.

Box 1. Comparator Programmes

We have identified several programmes, mostly ongoing, in Ghana that have similar aims to the MVP. We note them especially if they have a long horizon and there are indications that the programme will be evaluated. Also noted are a sample of programmes directed at affecting the MDG goals. During the baseline period we will further explore the feasibility of accessing and using outcome and cost data from these programmes.

1. Cash Transfer, Health Insurance and Health Outcomes in Ghana: LEAP provides cash transfers to ultra-poor households with orphans and vulnerable children, the elderly and disabled in Ghana. LEAP is funded by DFID as well and is now evaluated by ISSER and the University of North Carolina with 3ie funding. This is likely to be our primary comparator.

2. GoG Northern Region Poverty Reduction Programme (NORPREP) (2004-2012): NORPREP is a poverty reduction programme in northern Ghana.

3. World Bank Sustainable Rural Water & Sanitation Service: The objective of the Sustainable Rural Water and Sanitation Project is to expand access to, and ensure sustainable water supply and sanitation services in rural and small town communities in six regions of Ghana.

4. United States Agency for International Development (USAID): The Indoor Residual Spraying (IRS) project provides strategic, technical, management, and operations support in IRS for malaria control in targeted districts.

5. USAID Support for Malaria Program: To reduce the malaria disease burden in two regions.

6. DFID Prevention of Malaria through Procurement and Distribution of Long Lasting Insecticide Treated Nets.

7. DFID Market Development in the North (of Ghana): The primary purpose of this project is to expand market access for the poor and increase trade in the Northern Savannah.

8. Canadian International Development Agency (CIDA) assistance to Ghanaian Food-insecure Households in northern Ghana: The project supports the GoG in its efforts to improve the protection of vulnerable populations in northern Ghana from the devastating impacts of floods and droughts, and also to ensure that the immediate food needs of the most severely food-insecure households are met.

9. J-PAL Village Savings and Loan Association in Ghana: Researchers are working to measure the dynamics of self-selection with Village Saving and Loans Associations (VSLAs). This study is conducted with 180 communities selected by partner organisation, CARE, which identified them as villages in which they could initiate VSLA programmes.

10. J-PAL Targeting the Ultra Poor: The study is being conducted in grey communities in which other NGOs do not have a significant presence. Eligible households are identified with a Participatory Wealth Ranking (PWR) during which villagers rank the economic status of community members. Similar project being funded through 3ie with different researchers.

181. We will carry out a systematic search[33] for all related programmes in northern Ghana to establish links to share findings of these programmes at the earliest convenience for outside programme implementers. We realise the sensitivities involved in such interactions. Many Conditional Cash Transfer (CCT) programmes have multi-tier demand incentives affecting a range of outcomes and also

[33] This will be a literature review undertaken in a systematic manner, not a Systematic Review.

targeting the MVP. Along with the CCT aspect, schools and health centres have been refurbished or workers have received incentive payments. We believe these wider programmes could potentially provide good comparators. It is preferable that the foci of these studies are programmes in West Africa and Sub-Saharan Africa.

Positive and negative externalities

182. The programme site may impose *both* positive and negative externalities in nearby control areas. Any CEA will incorporate the cost of projects in the control area as well as the MVP area. It may be that due to funding allocated to MVP that the control receives lower amounts of fiscal allocations. This fact does not affect the incremental CEA or CBA of the MVP. However, from a public finance perspective, diversion of funds may further limit scaling-up of the MVP. An externality originates from changes in the allocation of public and project expenditure by government and NGOs determined by the programme. It is difficult to predict if this externality is positive or negative. District assemblies and government are committed to invest in villages to match the project investment. If the public budget is fixed this implies a reduction in public expenditure in non-MV localities. On the other hand, many public investments were planned long ago in the villages and therefore the displacement effect is likely to be less than the size of the investment requested by the MV project. Additionally, in the future other public or private initiatives may prioritise non-MV against MV areas because the area is already privileged by the MV project.

183. In order to detect these effects we have included two additional modules in the community questionnaire. The first module collects data on projects by any funding body in each locality, while the other module collects data on district assembly projects and expenditures on each of the projects and control localities. This information will be complemented by qualitative work directed to ascertain how public expenditure decisions at the district level are made, and on what is the likely size of expenditure displacement (as briefly discussed in the next chapter). If a rate of return to public investments can be calculated, a correction to the DD overall impact attributable to the MV interventions can be performed after accounting for public-expenditure related externalities (Chen, Mu et al. 2009).[34] However, the population covered by the intervention is 12% of the total population in Builsa and 9% of the total population in West Mamprusi. This suggests that the displacement of funds by public and private projects produced by the MVP should not be too large.

3.3 Risks and mitigation measures

184. The assessment of cost-effectiveness involves many challenges. This is in part due to the complexities arising from our attempt to compare the expected synergistic effects of the MVP with appropriate comparators, as well as in trying to gain access to reliable cost data. It is only once the evaluation team has had access to the accounting system of the MVP, has been able to fully assess how the Earth Institute (EI) intends to capture government and NGO expenditure in the MVP area, and gained access to data from potential comparator programmes, that the full extent of what is feasible will be known. This work will be undertaken during the baseline period, during which we propose the following activities as we work towards developing a protocol for the study:

- Identifying different components of the MVP
- Identifying modes of implementation, and our ability to isolate total costs of specific components
- Agreeing upon key outcomes of interests
- Developing a method to identify the relation between inputs and outputs which produce the outcomes. This may involve following costing methods adopted for specific types of programmes, for example, those for health or agricultural extension programmes.

[34] Chen S, R Mu and M Ravallion, 2009. Are there lasting effects of aid to poor areas? *Journal of Public Economics*, 512-328.

- Categorising inputs to fit with budget line items and attaching actual budget to inputs
- Developing a method to carry out costing exercises systematically on an on-going basis (e.g. quarterly)
- Identifying comparators by taking inventory of all projects that are being carried out in the nearby area and consulting to carry out a common costing methodology
- Consulting government accountants about how costing is carried out in the control areas

185. At this stage of the methodological design, it is acknowledged that there are a series of unknowns and risks for the evaluation in terms of measuring cost-effectiveness. The table below summarises the main risks, and our mitigation measures.

Table 11: Risks and mitigation measures for the cost-effectiveness study

Key risks	Rating	Mitigation measures
1. An inability to gain access to detailed SADA/ Millennium Promise accounting/expenditure data for the implementation of MVP.	High	We will work with the Earth Institute/SADA/Millennium Promise to develop a protocol/reporting format for receiving expenditure data on a regular basis. We will involve DFID-Ghana (who is funding MVP), if necessary.
2. MVP expenditure data is not linked to project activities or locations.	Low	We know that expenditure data is linked to activities to some extent, but have (to date) been unable to access detailed data to ascertain how this is linked, and whether it is possible to identify expenditure in particular locations within the MVP site. We may have to adjust the methodology, depending on the quality of the data available.
3. MVP attracts resources (financial and in-kind) from a variety of donors/organisations, which are not properly costed and tracked.	Medium	We understand that for each 'donation' to the MVP, a Memorandum of Understanding is signed. On-budget resources should be recorded in the MVP accounting system, and we will verify what is/ is not captured by comparing with the signed MoUs. We may need to use follow-up interviews with partner organisations to ascertain the detailed cost data.
4. Expenditure data on government or NGO projects in the area is not available.	Medium	We understand that the District Government records and collects on-budget data of projects in their area, and that EI intends to collect this data as well as off-budget data. We will assess the quality and availability of these datasets, and if there are gaps, we may collect the data ourselves – and this may in any case be required in the control sites.
5. Comparator data is unavailable or unsuitable.	Medium	Our exploratory visits suggest that such data should be available for at least the LEAP programme. Depending on what is available we will have to adjust the methodology accordingly.

4 Qualitative research and reality checks

186. Rigorous impact evaluations must also include an interpretation of *how* change has occurred as well as quantifying 'real effects' and attribution. Therefore, we include in our methodology a supporting qualitative module, which will provide, in the first instance, an additional *interpretational lens* on the quantitative findings, and will be sequenced after the quantitative surveys to take initial findings back to communities to obtain their interpretation of them. This will explore particular findings and diagnostic factors including: poverty and vulnerability, the empowerment and voice of the disadvantaged and marginalised, changes in intra-household relations (and these relationships as drivers of change), institutional and governance factors (including state and traditional systems), localised adaptations and synergy effects of the package of agriculture, education, health, and rural infrastructure resources and services provided by the MVP.

187. We will use four key qualitative methods

- Poverty and vulnerability assessment
- Interpretational lens using Participatory Rural Appraisal (PRA) tools
- Institutional assessment
- Reality checks

188. Most importantly this module will take community members' own measures of project success as the starting point.

189. In addition to the above, further information will be collected from district and regional government representatives and other development partners in the MV districts and regions (including a control district) to assess how government resources and services were utilised inside and outside the project as well as before and after, in order to assess how they change as a result of the project. Another area for the qualitative study will be on the relationship between the MVP and institutions (both government and non-government) in these districts and regions and changes over the period of the project. This may be linked especially to the cost effectiveness analysis within the evaluation if there are negative or positive externalities (see Box 2, below). Depending on the depth required, this could perhaps be part of an extended institutional assessment, or a separate survey module that would be added to the quantitative baselines.

Box 2. Capturing human and financial resource displacement

As mentioned in the section on externalities (Chapter 2), the analysis will take into account potential displacements of funds at the district level. These displacements can negatively affect control villages and can bias the DD estimates of programme impact. In order to assess these effects we have included two sections in the village questionnaire that collect data on existing projects by governments and NGOs and district assembly expenditures in human and physical infrastructure. However, a questionnaire pre-test revealed that the information collected in this way is both unreliable (village authorities are often unaware of projects and committed expenditures) and insufficient (it is the district level resources allocation that is needed, not just the repartition between project and control villages).

Hence, we will design a tool for collecting data on staff and expenditure at the district level on a yearly basis. Since we do not currently know how expenditures are allocated and the level of district discretionality employed, we decided to conduct some prior qualitative work. This work will consist of interviews with key people in Accra and local authorities and has the goal of identifying the best sources of information on committed expenditures and allocations. The same interviews will serve the purpose of ascertaining to what extent the programme has the potential to change current commitments and future resource allocations.

Objectives

190. The main thrust of the qualitative work will be to complement the quantitative work, which shows what has (or has not) changed, by shedding light on the reasons *why* and *how* changed has occurred, in

particular, from the perspective of community members with a special focus on the experience of disadvantaged and marginalised groups thereby responding to the key evaluation question: "*Does the MV package empower disadvantaged or marginalised groups (e.g. females, the disabled, or the elderly)?*" (ToR, Section 4.2 (a)).

191. The qualitative module will focus on addressing the following objectives:

- The need to **understand the reason why** – why a change in behaviour, why a statistical difference, why no change despite careful and deliberate intervention, why a spill-over into other communities. The qualitative survey is thus an opportunity to provide explanations through discussion in groups within communities or interviews with key individuals to drill down further on specific areas that arise from the analysis.
- The need to **complete the whole picture by addressing the gaps** that quantitative data typically cannot deal with – e.g. how programme interventions have linked or not, how relationships have changed at HH, community, district and regional levels, and among institutions, and people's views, perspectives, opinions and reasoning.

The field survey guide for the qualitative work will also weave the **enablers of and blockers to change** into the line of questioning in order to support lesson-learning on what is perceived by community members to work/not work and why.

4.1 Fit with the quantitative approach

192. The qualitative module will provide an additional interpretational lens on the quantitative findings at HH, community, district, and to a lesser extent, regional level. The essence of the module is that it will begin with baseline work that will help to determine some parameters/domains of change that need to be measured from a beneficiary perspective. Qualitative surveys will then be sequenced *after* the quantitative surveys, taking the initial findings back to communities to obtain their interpretation and also track change in relation to what community members see as truly beneficial. Our other main qualitative tool - Reality Checks - will provide an opportunity to understand the reality from the HH level. Both will pick up on any unexpected impacts.

4.2 Fit with the MVP process evaluation

193. The MVP Process Evaluation (PE) is a largely qualitative component of the MVP's M&E platform. The PE was initiated with the broad aim of working with site teams and key stakeholders to describe the development and implementation of the MVP programme and interventions. Informants have the opportunity to share their experiences as to how the MVP model was adapted to the local context; the sequencing, timing and rationale behind MVP implementation; and how and why the MVP model was customised for each site. In addition, the PE investigates the barriers and facilitators related to the project site's successful and unsuccessful programmes and interventions. Its key starting point is for the project to use it internally to help manage, improve implementation or report on progress. It is done internally, by the project staff, as part of implementation.

194. On the other hand, the qualitative module of the impact evaluation will be independently done by outsiders, making use of secondary data or information such as that from the MVP's process evaluation where necessary, as well as generating primary data/information. The starting point of the qualitative module, as part of the impact evaluation, is the community members' measures of change and interests. Just as with all impact evaluations, it is to be used externally for funding decisions, accountability reporting, policy, planning and budget decisions.

195. The qualitative data collected during the PE is used to enrich the quantitative data gathered by the M&E team from HH and female quantitative surveys. These combined data sources will provide insights and lessons for replicability, scale-up and transfer of the MVP model to other contexts.

196. Sector-specific questionnaires for focus groups and individual interviews have been developed for three levels of stakeholders: 1) MVP project staff and field implementers, 2) community and 3) government partners.

197. In terms of the community Focus Group Discussions (FGDs) the aim is to understand people's experience and opinion of the MVP with respect to the following:

- Targeting: How has the MVP selected beneficiaries? How were vulnerable groups targeted?
- Procedures: What happened and how was the communication between you and the MVP?
- Expectations: What did the community member expect from the MVP?
- Lessons learned: What could be done differently if we started over?
- Sustainability: Which interventions will last after the MVP leaves and why?

198. MVP has indicated plans to undertake a process baseline study in addition to its PE work. As much as possible we will co-ordinate with MVP with the aim of synchronising our work to theirs and also draw on the findings from the PE work and data gathering. We are mindful of the PE work in the design of our approach to the qualitative research and have taken it into consideration in the selection of methods, as well as key themes that we will focus on during this component of the evaluation. We anticipate that the two angles, one seeking to learn about project impact on the beneficiary (and the pros and cons of various aspects of the implementation approaches) and the other with the beneficiary perspective of the benefits and experience of the project, will provide a basis for comparison.

4.3 Key themes of the qualitative research

199. MVP is a complex set of inter-connected and multi-layered interventions which generates a multiplicity of different evaluation questions and themes that could be explored within the framework of the qualitative research. In order to ensure a focused approach to the qualitative component we have defined a number of core themes. These themes/strands will be the key areas explored by the qualitative instruments:

- Overview of the main features of the community (ethnicity, religious beliefs, livelihoods, recent notable trends, other external interventions, etc.). These will be captured in **village narratives** and will form part of the situational analysis work we envisage during the baseline.
- The qualitative survey assumes a multi-dimensional definition of poverty. However, how people define and experience poverty varies from place to place and for the different socio-economic groups. For this reason, for the qualitative study, the starting point will be community views on **wealth and wellbeing** and how these have changed over the duration of the MVP. Additionally, the qualitative study will get community members to interpret the findings from the quantitative study regarding the predetermined poverty indicators and other empirical analyses that impact evaluation is interested in testing against the big push theory of change.
- **Poverty and vulnerability** and the **coping strategies** people develop to respond to poverty and vulnerability, including migration and child poverty.
- **Empowerment and voice of disadvantaged and marginalised groups** as both an outcome and an explanatory factor.[35] This will include exploration of gender and age factors and construction of social capital and changes in intra-household relations and these relationships as drivers of change.
- **Change processes** in terms of engendering changes in knowledge, attitudes and behaviours of the community and its members.

[35] This section responds to the Evaluation Questions on: "*Does the MV package empower disadvantaged or marginalised groups (e.g. females, the disabled, or the elderly)?*" (ToR, Section 4.2 (a)).

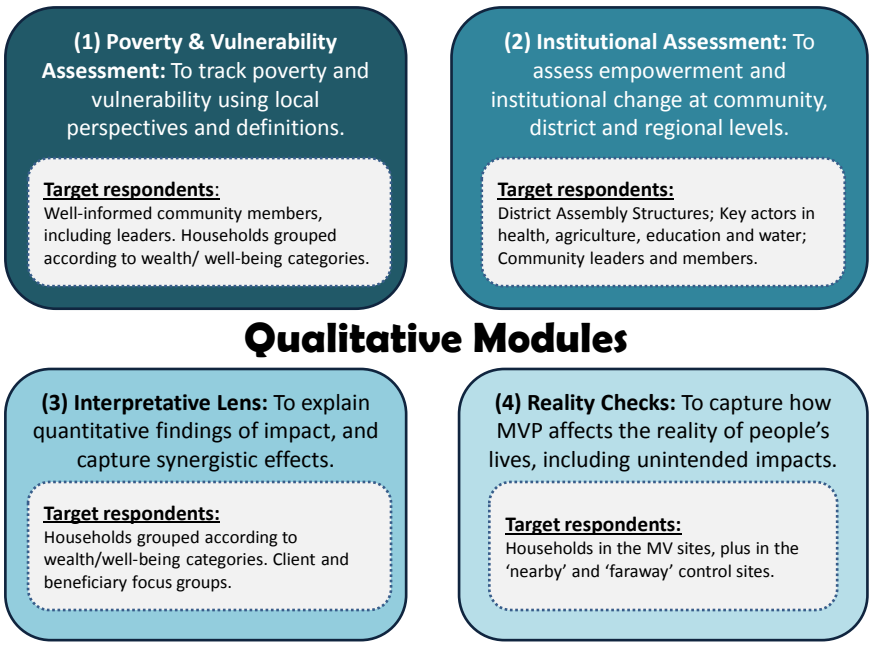
- **Institutional and governance factors** (including state and traditional systems) at community and district levels – how relationships between institutions, organisations and groups have changed as a result of the project and why. Impacts of the MVP on the culture of these relationships. Is the MV top down or bottom up? How has it impacted on institutions and organisations that serve the community or district?
- The synergy effect - how are the package of interventions working together and generating more value to the community?

4.4 Methodology for data collection

200. We will use four main research methods in our qualitative research. Figure 12 presents an overview of these four methods, the main target groups and the key objectives of the research method.

- **Poverty and vulnerability assessment** (PVA) to track and assess how the interventions are addressing poverty and vulnerability and how they are impacting on migration and child poverty.
- **An interpretive lens of the quantitative findings** (drawing on PDA's methodology used for the Participatory Poverty and Vulnerability Assessment (PPVA) work in Ghana).
- **Institutional assessment** to assess empowerment, voice and relationships at community and district levels.
- **Reality Checks** to provide insight on how the MVP is experienced at the HH level.

Figure 12: Overview of Research Methods



201. The qualitative work will take place in three main stages: Baseline and then Year Three (mid-point) and Year Five (end-point).

202. During the baseline, we will seek to establish the criteria people use to define poverty and to measure wealth and wellbeing. Since the goal of the MVP is to reduce poverty, it is important to know from the beginning what measuring sticks community members themselves use to interpret results as they

come. These will help to track impact on the most marginalised and vulnerable members of the community, and determine whether the interventions are correctly targeting these people/groups.

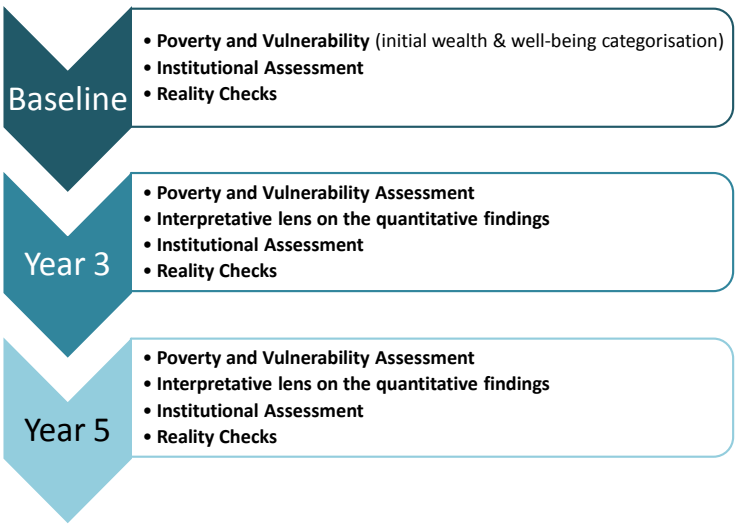
203. A key activity during the baseline is the situational analysis which initiates the tracking of poverty and vulnerability and examines how the MVP over time addresses the causes of poverty and builds people's resilience, especially poor HHs. This work will include the development of a portfolio of village narratives that will be two-to-three page overviews of the main features of each community. These will outline the livelihoods of men and women, key social and political institutions, key social and economic services available to the community, ethnicity, recent and notable trends, other external interventions going on in the community, etc. Some of this information will be taken from the PE, the quantitative and qualitative baseline studies, and existing research and literature of the project districts.

204. Another key activity during the baseline is the institutional assessment at community, district, and to a limited extent, of SADA Secretariat. A project like the MVP that is multi-dimensional, multi-layered and interconnected is bound to impact on existing institutions and organisations. It will also give rise to new ones. This dimension of the qualitative research will be to capture the existing relationships between institutions, organisations and groups at both community and district levels, how these have changed as a result of the project and why.

205. The qualitative work in years three and five will be sequenced after the quantitative work in order to check back with the community members on quantitative findings, many of which will relate to the interventions in key sectors – agricultural, health, education and water. But they will also revisit the key strands as listed above in Chapter 3.5. The research team will put in place mechanisms for co-ordination that will enable the qualitative research team to identify the findings from the quantitative work that need to be drilled down on. The quantitative research may identify communities in which more or less progress is being made, as well as successful and less successful interventions e.g. people not using/purchasing fertiliser. The qualitative will pick on such findings for further interrogation. Similarly, the quantitative data will be modified in the mid- or end-terms to collect data on new leads from the qualitative data. To make this possible, for each round, an internal team meeting/workshop in Ghana will be held. This will be used to discuss preliminary findings from quantitative and qualitative methods, emerging issues from control/treatment sites, and areas for further investigation.

206. An overview of the planned research during the course of the impact evaluation of MV is provided in Figure 13 below:

Figure 13: Overview of qualitative research at baseline, mid-term and final evaluation stages



207. Further detail on the PVA, the institutional assessment and the interpretative lens and the key research questions which will be explored by each tool is provided in Appendix E. Information on the Reality Check methodology is also contained in Appendix E. As outlined in Appendix E and the third column of the issues and question table the research team anticipates using the following methods:

- **Wealth/wellbeing ranking** in discussing poverty and vulnerability, starting with the communities' own criteria and categorisation and getting them to assess whether the MVP's 'big push' theory of change has made any difference to their wealth/wellbeing by both the communities' own criteria and that set by MVP.
- **Seasonal mapping** for seasonal variations and their impact on the poor and vulnerable groups like women, children, ethnic minorities, and whether these factors were taken on board by the MVP in the delivery of its interventions.
- **Trend analysis** for changes over time to assess whether MVP has halted some negative trends, for instance of migration, environmental decline, or in education, health, infrastructure, etc. or resulted in some negative changes.
- **Ranking and scoring** to get a sense of proportions or for prioritisation by the poor in terms of importance, relevance, change, etc. and proportions.
- **Mobility mapping** on the issue of migration, since it is a major coping strategy of the poor and vulnerable like women and the youth it will be important to see the impact MVP is making on this.
- **Income and expenditure analysis** for cost-benefit analysis of new technologies in farm and off-farm activities.
- **Institutional mapping** for analysing the level of community involvement, esp. of the poor and vulnerable and of the institutions that matter to them, e.g. of their own self-help groups, and analysis of local government and community ownership of MVP interventions, the impact on local government and community institutions and how the involvement of the poor and vulnerable in governance in general and the MVP in particular have changed over the five years.

208. The training of the field teams involves a field test that usually enables the team to fine-tune the methods so that there is consistency across the field teams. An overview of each of the evaluation tools is provided below in Tables 12 to 15.

Table 12: Qualitative toolkit: poverty and vulnerability assessment

Tool	Poverty and vulnerability assessment
Main aim or purpose	To track poverty and vulnerability of households to see if the interventions are addressing the causes of poverty and building people's resilience, especially those of poor households, using community's own criteria of wealth or wellbeing.
Approach	The theory of change behind the poverty and vulnerability assessment is that by its design, the MVP assumes a multi-dimensional definition of poverty. It therefore seeks to address the manifestations of poverty, such as low income, food insecurity, low level of service provision resulting in high maternal and child mortality, low educational outcomes, as well as powerlessness (especially of women), etc., through an integrated, multi-sectoral approach. For this reason, the extent to which poverty is reduced in the MVP communities and has set them on a sustained path of growth should be the final outcome of the total sum of the MVP interventions. This outcome should not be measured based only on some predetermined indicators but also from the perspective of the poor themselves, using their own criteria or indicators. In this regard, a small sub-set of community members (village heads, sub-village heads, family heads, leaders of women and youth groups, etc.) who know the households in the community and their wealth/wellbeing levels will generate criteria for wealth or wellbeing in their community and use the criteria to categorised a randomly selected number of households into different wealth/wellbeing groupings (e.g. 'rich', 'average', 'poor' and 'very poor', etc.). The focus groups that will be formed from the household categorisation will then assess the impact of the MVP on poverty and vulnerability and the targeting of the interventions based on the community's criteria for wealth/wellbeing.
Target Groups including any specific groups	Households in the MVP and comparison/control communities, ethnic minorities, persons with disabilities (if not captured in the wellbeing categories), children, the aged.
Numbers to be consulted/engaged in the process	By its nature and purpose, the qualitative research will not be conducted in all the MVP and control communities. The team is proposing that it be done in 20 communities – 10 MVP communities and 10 control communities. Of the control communities five will be near communities and five far communities. In each community, 8-12 well-informed community members including leaders will be selected to undertake the generation of criteria and the categorisation of households into wealth/wellbeing categories. Between 32 and 48 people from four wellbeing focus groups, 16-24 children from two children focus groups, and probably another 16-24 from focus groups of ethnic minority or aged.
Method of selection of participants	Depending on the number of households in a community, 60-100 households would be randomly selected for the wealth/wellbeing categorisation. This is to allow for a sizeable number from which to form at least four focus groups of 8-12 people each. These households may or may not include households in the quantitative survey. From the wealth/wellbeing categorisation, at least four focus groups to be formed taking the following into consideration: • A wealthy/average male group drawing names from each category • A wealthy/average female group drawing names from each category • A poor/very poor male group drawing names from each category • A poor/very poor female group drawing names from each category This however, would depend on the categories the community members come up with. The focus groups from the wealth/wellbeing categorisation additional focus groups of children, ethnic minorities and aged will assess over the five years, the complementarities between the package of interventions, how they are contributing to a sustained growth path, as intermediate outcome measures, the targeting of these interventions, and how they impact finally on poverty, vulnerability and coping strategies (especially migration and child poverty), as defined and experienced by the different community members.

Table 13: Qualitative toolkit: an interpretive lens of the quantitative findings

Tool	An interpretive lens of the quantitative findings
Main aim or purpose	To assess the synergy effects - how the package of interventions are working together and generating more value to the community.
Approach	The theory of change behind the interpretive lens of the quantitative findings is that the activities, outputs and outcomes of the MVP-SADA seek to deliver a co-ordinated 'package of science-based proven interventions for agriculture, education, health, and rural infrastructure with the leadership of strengthened local governments and communities'. It would be inefficient and ineffective for the qualitative study to seek to assess the delivery, adoption, impact and sustainability of the numerous individual interventions of the MVP. Instead, the team will select appropriate PRA tools to provide an additional *interpretational lens* on the quantitative findings and that from the PE of the synergy effects and cross-sectoral implementation issues of the various interventions, namely: agriculture, education, health, gender, environment, energy/Information Communication Technology (ICT) interventions, water and sanitation. In this vein, the qualitative survey will explore positive and negative changes in the following: • Asset base, especially productive water (e.g. dams, irrigation) that reduces communities dependence on rain-fed agriculture • Markets for goods and services • Agricultural inputs and services (e.g. vet, extension services, fertiliser), their delivery and adoption • Access to financial services • Expanded employment opportunities • The absorptive capacity and or uptake of the facilities and services provide by MV interventions • In-migration into MV communities from other communities to benefit from the project It will therefore be sequenced after these surveys/studies, and will also make use of existing researches or literature on these issues.
Target Groups including any specific groups	The same people in the focus groups from the wealth/wellbeing categorisation above, and additional focus groups of children, ethnic minority, youth, aged, etc. will be asked about their perceptions and assessment of these interventions. Beneficiary/client groups – where these are different from the focus group members, additional beneficiary/client focus groups will be formed.
Numbers to be consulted/engaged in the process	Same number of communities as proposed above. In each community, between 64-96 people from the wellbeing, ethnic minority, children, aged focus groups and another 40-60 people from five beneficiary/client focus groups and key individual interviews.
Method of selection of participants	The four wealth/wellbeing focus groups, and ethnic minority, children and purposeful selection of institutional actors and client groups based on their first hand knowledge/experience of the issues emerging from the quantitative evaluation.

Table 14: Qualitative toolkit: institutional assessment

Tool	Institutional Assessment
Main aim or purpose	To assess the empowerment and voice of disadvantaged and marginalised groups (especially within the context of inter and intra-households relationships), how the relationships between institutions, at both community, district and regional levels, have changed from the perspective of both the community members and the key actors in the institutions over the course of the project, and to interrogate any governance issues that may arise from the quantitative survey.
Approach	The theory of change behind the institutional assessment is that a project, like the MVP, needs to work with and strengthen the institutions that exist and these should take the lead in making change happen. (Only when necessary should new organisations or institutions be created). If they are weak their capacity needs to be strengthened to be able to undertake and sustain the change being sought by the project. A project like the MVP is temporary and thus needs to consider its legacy from the outset and in its design, in terms of the impact it has on the institutions that remain, in particular, in terms of enabling them to lead their own development. This implies ownership and inclusion, not just participation, by the people themselves and their institutions. For instance, it will be important to know how the poor and marginalised have been brought into decision making processes so that their needs and concerns are addressed not just during the project but beyond. The main actors in various institutions (see row below) will be interviewed about the impacts of the project on them (how their capacity has been built/strengthened or weakened and how the MVP has positioned them to sustain the outcomes of the project post-implementation as it is these institutions that will sustain the work into the future. The relationships between these institutions and actors and how these change will also be examined. Reference will be made to existing studies or reports of community and district-based projects and their impact on institutional actors. In addition, community members, the same people in the focus groups from the wealth/wellbeing categorisation (ethnic minority, children, aged) will be asked about their perceptions and assessment of these institutions, inter and intra-household relationships. Additional client focus groups will be formed. For example, the parents in the first focus groups may be asked about the school but their perspective will not be the same as the students. In year three and five – these same groups and lines of enquiry will be revisited in order to track thematic trends and changes in both content and in perception. We expect this to differ from the PE as it will focus on intended and unintended impacts, with community member experience at the centre, rather than concerns about the effectiveness of implementation strategies. As well as deliberating on state organisations, such as service delivery agencies, governance institutions, and traditional authorities, informants will also be asked to look at inter and intra household relationships. How the project has affected gender relations, family ties etc., in-migration from other communities. As the Reality Check will focus intensively on the HH level this community group perspective with be reviewed alongside this and the PE outcomes to triangulate the information.
Target Groups including any specific groups	1) Main actors in institutions: • *Staff/actors of district assembly structures* – district assembly, area council, unit committees/assembly persons • *Key actors in health, agriculture, education and water (at regional, district and community levels)* – which activities of the programme are providing more synergy • *Any new institutions created by project*. E.g. a committee to manage a dam, co-operatives. etc. • *Traditional authorities* - both chiefs and elders and women leaders (magazia) 2) Community members - The same people in the focus groups from the wealth/wellbeing categorisation, ethnic minority, children, aged above will be asked about their perceptions and assessment of the institutions. 3) Beneficiary/client groups – where these are different from the focus group members, additional client

	focus groups will be formed. For example, the parents in the first focus groups may be asked about the school but their perspective will not be the same as the students.
Numbers to be consulted/engaged in the process	Same number of communities as proposed above. In each community, between 64-96 people from the wellbeing, ethnic minority, children, aged focus groups and another 40-60 people from five beneficiary/client focus groups and key individual interviews.
Method of selection of participants	The four wealth/wellbeing focus groups, and additional focus groups from ethnic minority, children, aged and purposeful selection of institutional actors and client groups based on their first hand knowledge/experience of the issues emerging from the quantitative evaluation.

Table 15: Qualitative Toolkit: Reality Checks

Tool	Reality Checks
Main aim or purpose	Mini anthropological-style study to better understand how MVP affects ordinary people in reality at the individual and household level.
Approach	There are domains of change that RC seeks to pick up on are: • How intra-household relationships (gender and generational) are affected by the MV and other interventions • How inter-household relationships are affected (especially between the poor and non-poor) The Reality Check approach is one of immersion, involving the collection of insights to assess the impact of MVP by directly gathering the experiences, opinions and perceptions of people living in the MV. Researchers stay for about four days and nights in the homes of poor families and join in their lives. They engage in unstructured conversations, having been briefed in the areas of interest to the project. The Reality Checks will take place ahead of the PRA study so that it can help to highlight particular issues that the PRA needs to follow up on. This will help the PRA team to adjust or fine-tune the content of their semi-structured interview guide. The Reality Checks and PRA team leaders will work in tandem to ensure there is complementarity in their work – thus optimising triangulation, coverage of as many issues as possible and avoiding duplication. The Reality Checks will take place in different communities from the PRA, as past experience has shown that the Reality Checks team needs to be as free of participant assumptions and expectations as possible. The lead researcher will take part in the immersion in the pilot period as well as the first round thus enabling supervision, her own analysis and debriefing of both teams during the field period.
Target Groups including any specific groups	The Reality Checks will focus on HH in both the MVP area itself and both near and far control areas. These will be selected to be representative of the different wealth categories identified during the PRA work as far as possible as well as ethnic difference where appropriate. Within the HH particular attention will be given to typically marginalised groups and individuals as well as migrants.
Numbers to be consulted/engaged in the process	A team of seven researchers (two teams of three researchers plus a team leader/lead writer) will undertake the Reality Checks. It has been designed for the researchers to work in teams, as against a single researcher per village for longer, in order to reduce the problem of researcher bias. The way it is organised with several researchers in the area, but within their own immediate locations, provides a way to mitigate this. The teams will undertake two rounds of immersion as well as the pilot (training) immersion, thus the two teams will go to three communities each and each team member will stay in a different household. The team plan to cover 19 households (13 in MV communities and 6[36] in control communities) during a period of 22 days. These 19 households will be in six different communities and they will be selected roughly on the basis of representation from different districts/regions/ethnic groups, as well as practical considerations for the researchers such as distance, though these factors will be secondary if necessary.
Method of selection of participants	HHs will be selected by the researchers through discussion with community members. We will seek to revisit the same households at baseline, mid-term and final evaluation stages in order to track change over time.

[36] Three HH in the near CV, three HH in the far away CV.

4.5 Sampling of communities of the qualitative research

209. Considering that the primary objective of the qualitative impact evaluation is to complement the quantitative research in finding answers or explanations for expected and unexpected changes or anomalies, the team has in mind to use purposive sampling to select 20 communities for the study. The sampling criteria will seek to capture the diversity of the communities in terms of demography (large and small communities), ethnic and linguistic diversity, including minority groups like the Fulanis who are also nomadic, accessible and hard to reach communities, better served and underserved communities, regional and district mix (old and new districts). The purposive sampling will also take into account other factors that the MVP process M&E and quantitative data will throw up like, for example, communities that are doing well and those that are not doing well. Using these criteria, 20 communities for the qualitative research will be decided with MVP team in the SADA zone.

4.6 Analysis and reporting

210. Within the qualitative research there will be four layers of analysis and reporting:[37]

- Taking **field notes**: Intrinsic to the qualitative survey is some level of community analysis. The poverty, vulnerability, migration and institutional assessments will all be that of community members. In addition, in undertaking the 'interpretive lens' of the quantitative findings, people in the different focus groups from the wealth/wellbeing categorisation will be analysing and assessing the MVP package of interventions.

  Detailed and accurate recording of all the discussions and the visual outputs (PRA diagrams etc.) is therefore of crucial importance in the reporting process. Given the huge quantum of information and analysis that is generated during a qualitative survey, it is very easy to lose and forget a lot of it, if it is not recorded immediately in the field. This is the basic data that can be used for further analysis and synthesis. It is for this reason that the role of the documenter is very important in the team.

- **Daily reviews**: After completing the fieldwork for the day the researchers will meet to reflect on the day's process and share their experiences with each other. Daily reviews are important, especially when the facilitators are divided into different teams and work with separate groups from the community. This review makes it easy to triangulate and analyse the results.

  The daily review also helps in reflecting on the progress made and in planning for the next day's fieldwork. Information that needs to be triangulated is identified and issues not explored are then included in the next day's plan.

- Writing the **synthesised site report** for every community/site where the study is conducted: Before this report is written it is necessary for all the facilitators to **review the data and process** together. All field notes are analysed before conclusions are reached. The checklist of issues used for the fieldwork is revisited. All the information available on each of the topics is collated and analysed. Any new themes or topics that may have emerged during the study, and not listed in the checklist, will be added.

  It is quite common to get multiple, and sometimes contradictory, responses on a topic that do not match. These can depend on the level of diversity within the community. The site/community report will **reflect this diversity**. It will also clearly indicate **results that cut across the different socio-economic groups** within the community.

  In case there are any gaps in the information, or some questions remained unanswered, this will be clearly stated in the report.

[37] The system of reporting in relation to the Reality Checks is detailed in Appendix E part 2 – since this method requires a specific approach.

- Preparing an **overall field synthesis report**, based on the findings from all sites/communities: The team leaders and all the members of the research team will be brought together for a joint analysis workshop where, together, they will review the data/information gathered from all the communities. The Lead Writer then compiles a first draft that is reviewed by other team members. The process of writing may raise gaps in the data that have to be filled by returning to the respondents.

211. Steps (iii) and (iv) above will be done independently (in order not to be influenced) but at the same time it will be in relation to the sections/themes/areas in the quantitative survey as well as the parameters of the Reality Check. In this way the synthesis of the findings will begin early.

212. Following this the various teams of researchers (quantitative and qualitative) will meet to assemble a combined picture through joint analysis.

5 Independence and governance of the evaluation

213. This Chapter sets out the core principles agreed for the governance of the evaluation. It provides a brief overview of the governance structure, sets out how independence is to be achieved in practice, and also, the quality assurance/verification process to be implemented by the impact evaluation team.

5.1 Independence of the evaluation

214. An Evaluation Advisory Group (EAG) has been established to guide the strategic direction of the impact evaluation, reviewing and advising the DFID MV Adviser on decisions affecting the evaluation. The EAG includes representation from the ITAD consortium, MVP (including the Earth Institute), DFID, GoG, SADA, Local Government and others as per the EAG ToR. The EAG will meet regularly at quarterly intervals, and also before and after each major survey event. More regular communication will also be conducted by correspondence on specific issues as necessary. The EAG will report, via the chair or nominated group member, to the MVP Steering Committee and the MVP Stakeholder Consultative Group – both established in April 2012. The EAG has an advisory function but cannot approve or make decisions on behalf of the impact evaluation team. The prime purpose of the EAG is to provide constructive inputs from a range of stakeholders that will help ensure that the evaluation remains policy relevant.

215. A separate external PRG has been set up to formally peer review the initial design of the impact evaluation, and to review subsequent outputs from the evaluation team.[38] The PRG will review the scientific and technical quality of the evaluation, and in doing so, ensure that the design and implementation of the evaluation is robust and credible, and will stand up to external scrutiny. The PRG is viewed as an important part of the process for ensuring the independence of the evaluation.

216. The evaluation will conform to DFID's 'Ethics Principles for Research and Evaluation' which states:

"*Research and evaluation should usually be independent of those implementing an intervention or programme under study. Independence is very important for research and evaluation; in fact evaluations in DFID can only be classified as such where they are led independently. Involvement of stakeholders may be desirable so long as the objectivity of a study is not compromised and DFID is transparent about the roles played. Any potential conflicts of interest that might jeopardise the integrity of the methodology or the outputs of research/evaluation should be disclosed. If researchers/ evaluators or other stakeholders feel that undue pressure is being put on them by DFID officials, such that their independence has been breached, this should be reported to the Head of Profession for Evaluation who will take appropriate action.*"

217. The impact evaluation team and the MVP (particularly the Earth Institute) have agreed to work cooperatively as far as possible, although a number of 'red lines' have been established to ensure independence. These are set out in the paragraph that follows. As far as possible, the impact evaluation team will liaise with the Earth Institute, which is responsible for the M&E system rather than MV implementation. This will include issues such as the coordination of timings for all activities around data collection by the Earth Institute and additional data collected by the evaluation team, the coordination of meetings with communities and local government officials (to reduce unnecessary fatigue or confusion), and shared logistical and field support where appropriate.

218. The ITAD-led impact evaluation team will determine the design of the evaluation, in consultation with MVP through mechanisms such as the EAG. The independence of the design and conduct of the evaluation will be achieved by:[39]

[38] *Terms of Reference: Millennium Village in Northern Ghana*, Peer Review Group, February 2012, DFID-Ghana.
[39] As set out in the "*Principles governing the relationship between DFID, ITAD and the Millennium Villages Project (MVP) for the evaluation of the Millennium Village in Northern Ghana project*", March 2012, DFID-Ghana.

1. The impact evaluation team will determine the selection of non-intervention communities (the control) using the strategy set out in Chapter 2.3. Advice on these decisions will be sought through the EAG and PRG. There is an agreement with MVP to respect the impact evaluation's sampling frames.

2. Quantitative research. The impact evaluation team will sign-off (through the PRG) on the essential core content of the census and proposed baseline, mid and endline survey instruments to ensure that they meet the requirements of the impact evaluation design. The impact evaluation team may need to produce its own instruments as and when required, and the MVP may collect additional information for the purpose of informing programme activities.

3. The impact evaluation team will have the opportunity to input into the qualitative and economic costing components of the MV M&E system. While efforts will be made to harmonise with established MV systems in order to avoid duplication, the impact evaluation team will also conduct additional independent data collection where necessary. This is planned for the qualitative research component of the impact evaluation, as well as for costings.

4. MVP agrees to the auditing of survey data by ITAD or its representative during or following survey rounds. This will be used to check the quality and independence of data collection.

5. MVP agrees that the impact evaluation team will be provided sufficient time and opportunity to input on performance monitoring indicators, economic costing methods and other internal M&E activities being undertaken at the site. Also, that quarterly MVP performance data will be shared.

6. The Initial Design Document (IDD) will be formally peer reviewed by an independent PRG, made up of internationally respected experts. The selection of experts has been undertaken by 3ie, in consultation with MVP and ITAD. The selection has been on the basis that peer reviewers have a credible academic reputation, are independent and willing to constructively engage with the evaluation.

5.2 The MVP's quality control systems

219. This is a brief overview of the MVP quality control systems for capturing enumeration and data processing errors. There are three main parts to this system: (1) *Field based systems 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 a third time by a field supervisor); (2) *Random spot-checks* of enumerators are conducted by field supervisors (with protocols for the random survey spot-checks); and (3) 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).[40]

Protocol for fielding the HH survey quality control form

220. The **HH Survey Quality Control Form** is designed to verify the accuracy of survey data collected by enumerators. Supervisors will select a household, at random, for verification and will re-interview the main respondent, asking them a series of short, non-sensitive and easily verifiable questions (10-15 minutes), questions whose responses will be recorded, and then compared to the completed questionnaire previously filled out by enumerators.

221. **Selection process for verification:** During the first 10 days of enumeration, supervisors should randomly select at least one household per day for re-enumeration/verification visits, from the target list of households that will be enumerated that day. When selecting HH, supervisors should ensure that

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

every enumerator has had at least one revisit during the verification period. The same process should be repeated, during the last 10 days of enumeration.

222. **Preparation:** Once the supervisor has selected the household for re-enumeration, s/he should collect the questionnaires completed by the enumerators, and carefully fill out the fields highlighted in blue in the Quality Control from. They include: sub-village code, name of HH head, HH identification, and name of main respondent.

223. **Verification:** Once the supervisor has re-interviewed the HH and recorded the responses, s/he should compare the responses with those collected by the enumerator, and note any discrepancies in the KEY VALIDATION QUESTIONS of the Quality Control Form. The questionnaire and quality control form should then be handed to the data manager, so that s/he can also take stock of any discrepancies, and make corrections to the original questionnaire, as needed.

224. **Target Sample for Verification:** Supervisors should aim to select at least one household for per day for re-enumeration/verification visits during the first 10 days and the last 10 days of enumeration, for a minimum of 20 household questionnaires (about 5-10% of the original sample) for verification.

Protocol for fielding the adult female quality control form

225. The **Adult Female Quality Control Form** is designed to verify the accuracy of survey data collected by enumerators. Supervisors will select an individual respondent, at random, for verification and will re-interview the respondent, asking them a series of short, non-sensitive and easily verifiable questions (10-15 minutes) whose responses will be recorded, and then compared to the completed questionnaire previously filled out by enumerators.

226. **Selection Process for Verification:** During the first 10 days of enumeration, supervisors should randomly select at least two female respondents per day for re-enumeration/verification visits, from the target list of respondents that will be enumerated that day. When selecting respondents, supervisors should ensure that every enumerator has had at least one revisit during the verification period. The same process should be repeated, during the last 10 days of enumeration.

227. **Preparation:** Once the supervisor has selected the individuals for re-enumeration, s/he should collect the questionnaires completed by the enumerators, and carefully fill out the fields highlighted in blue in the Quality Control from. They include: sub-village code, name of household head, household ID, name of female respondent, respondent identification, and Q517.G (birth history).

228. **Verification:** Once the supervisor has re-interviewed the individual and recorded the responses, s/he should compare the responses with those collected by the enumerator, and note any discrepancies in the KEY VALIDATION QUESTIONS of the Quality Control Form. The questionnaire and quality control form should then be handed to the data manager, so that s/he can also take stock of any discrepancies, and make corrections to the original questionnaire, as needed.

229. **Target Sample for Verification:** Supervisors should aim to select at least two respondents per day for re-enumeration/verification visits during the first 10 days and the last 10 days of enumeration, for a total of at least 40 female adult questionnaires (constituting about 5-10% of the original sample) for verification.

Data entry and cleaning in CSPro

230. Questionnaires are single entered by a data clerk, cleaned, and then passed on to a different data clerk for double entry, and another round of cleaning. The data entry templates and cleaning scripts contain four basic types of checks: (i) missing data, (ii) invalid response codes, (iii) logical/consistency, and (iv) structural checks. An overview of the data entry and cleaning systems, using the education modules as an example, is provided below. Further details are set out in Appendix E.

Table 16: 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.

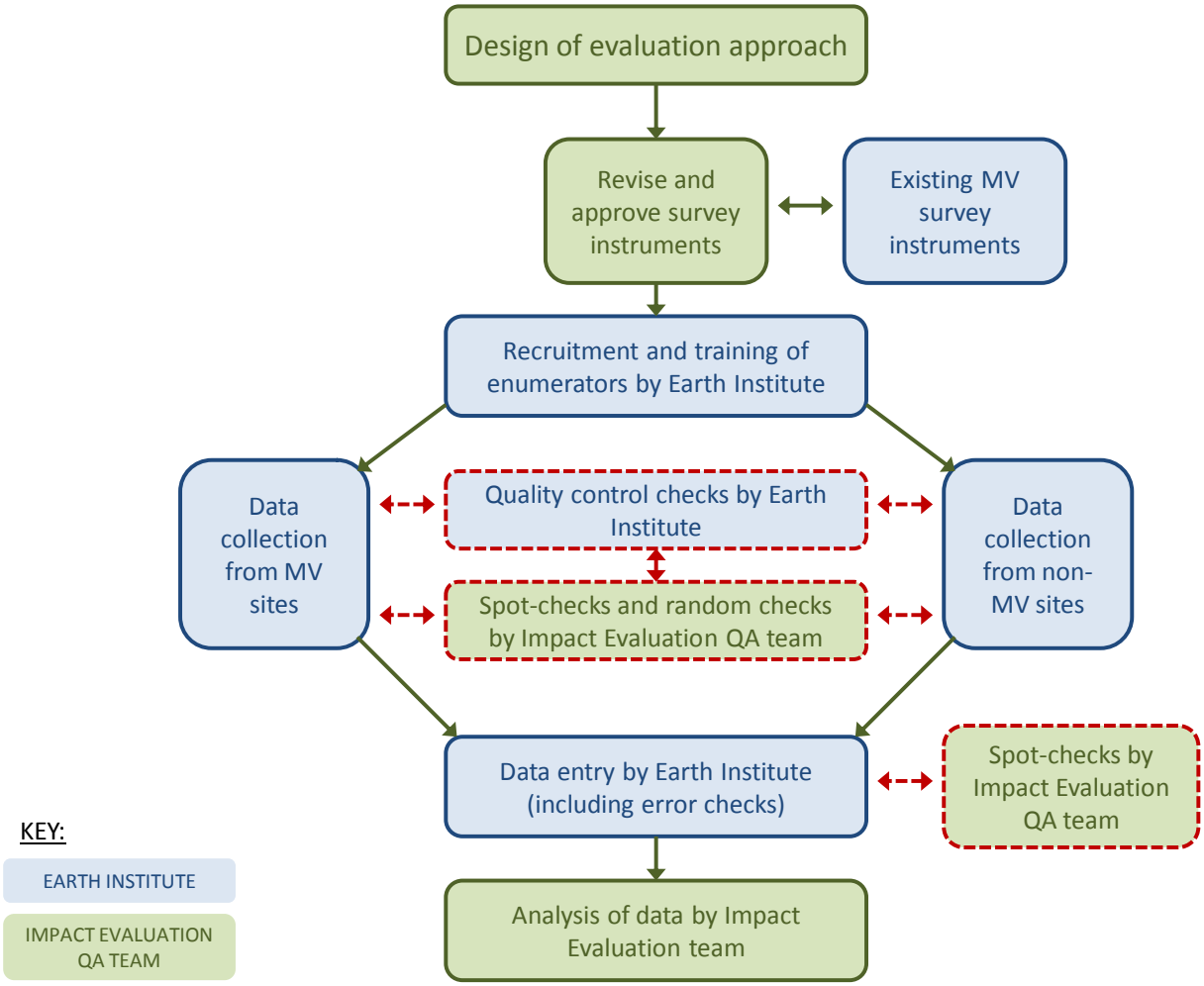
5.3 Verification and quality control by the impact evaluation

231. This Chapter sets out our processes for the quality assurance of the data collection and data entry, a critical role as these data processes are outside the aegis of the Impact Evaluation team. Since it is the MVP team, under the supervision of the Earth Institute, who will be implementing the quantitative survey instruments and collecting the data, the issue is doubly critical, given the controversy that has arisen over the MVP's assessments of its own performance. While we do not anticipate any issues over data manipulation,[41] the whole evaluation, from design, through data collection and entry, to analysis and reporting must be *seen* to be entirely rigorous and stand-up to public scrutiny. Hence we are deploying resources specifically to address the issue of credibility – reliable data, properly collected, using a robust survey instrument. The Chapter sets out our processes for ensuring the independent and transparent process for data collection and analysis.

232. The IE quality assurance and verification will take place at key stages in the data collection and analysis process. The focus will be on independently verifying the **implementation of the MVP's quality assurance systems** for each stage in the survey cycle, and, paying particular attention to assessing the prevalence of **research bias** (in particular, selection bias and expectation bias). Figure 14 sets out the division of responsibility between the Impact Evaluation Quality Assurance (QA) team and the Earth Institute's (responsible for the monitoring and evaluation of the MVP).

[41] While superficially having the MVP undertake the data collection may appear inherently biased, the data collection and analysis process is actually under the supervision of the Earth Institute (which is responsible for M&E, and *not* implementation). It is in any case difficult to see how individual enumerators could systematically manipulate the data, and instead reliability is more likely to be undermined by research biases.

Figure 14: Process of data collection and quality control



233. There are two types of errors associated with most forms of research: (i) **Random errors**, i.e., those due to sampling variability or measurement precision, occur in essentially all quantitative studies and can be minimised but not avoided; and, (ii) **Systematic errors, or *biases***, are reproducible inaccuracies that produce a consistently false pattern of differences between observed and true values. Both random and systematic errors can threaten the validity of any survey's results, although random errors can be easily determined and addressed using statistical analysis, whereas most systematic errors or biases cannot. This is because biases can arise from innumerable sources, including complex human factors.

234. The main types of bias and how these are to be addressed are set out in Appendix E.

The QA processes of the impact evaluation

235. In addition to MVP's own processes for quality assurance, the Impact Evaluation's own QA team will: (1) Independently verify the **effectiveness of the MVP quality assurance systems** in place for each stage in the survey cycle; and (2) Pay particular attention to assessing the prevalence of **research bias,** in particular, selection bias and expectation bias. The work of this team will focus on:

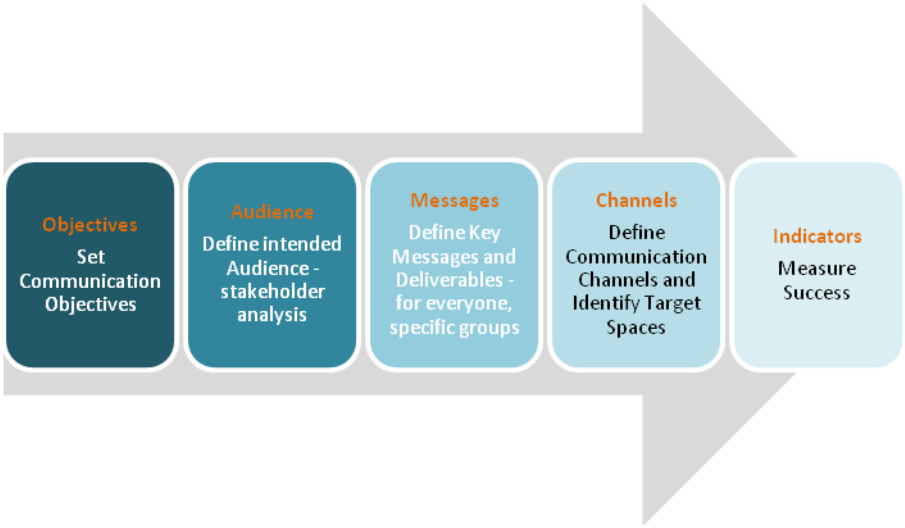
- **Field verification visits** to check the adherence to the MV quality control processes. This process will review whether procedures are being followed in the training of enumerators and by supervisors in the field. These checks include the shadowing of enumerators and supervisors during enumeration (random selection of 5% of the sample) to ensure adherence to the set requirements. The Process Check Tool is an overall comprehensive process checklist based on the survey planning and implementation Gantt chart which the QA team use to check documented MVP (Earth Institute) processes and procedures are being correctly followed.
- **Random spot checks of HH**. We will re-visit a sub-sample of households (random selection of 5% of the sample) and check that: (i) the HH members can be located and exist; and, (ii) check responses for a selection of questions within the survey instrument to test errors – as a result of expectation bias or measurement inaccuracy.
- **Statistical tests** to test the integrity of the survey data. We propose the application of *Benford's law* which is based on the peculiar observation that certain digits appear more frequently than others on datasets. We will also analyse the functional forms and the densities of key outcome indicators in order to test their similarity with commonly observed behaviours (for example, we expect the income density to be lognormal). In addition, we will screen values of the standard deviations of particular outcomes to check that they are in line with the values normally observed in similar studies (for example, a standard deviation above one for nutritional Z-scores is a sign of large measurement error).

236. The protocols for the verification visits and spot-checks are set out in Appendix E part 3.

6 Communication strategy

237. Good communication – in terms of messages, timing and the appropriate use of channels – is essential for the smooth-running of the MV evaluation and to support its integrity. The MV approach attracts considerable publicity, and there has been considerable methodological controversy around how to evaluate the MV in recent years.[42] Therefore, we view the Communication Strategy as not simply about disseminating evaluation outputs, but rather a more continuous two-way engagement throughout the project cycle, in order to reinforce a clear understanding of the project's goals, mitigate reputational risk and stimulate debate. In particular, the Communication Strategy aims to ensure engagement, transparency and accountability to project stakeholders, project partners and interested parties, stressing the value of conducting the impact evaluation and sharing results in an informative and timely manner. The strategy also aims to engage with researchers and the evaluation community, sharing methodological approaches and findings and supporting discussions and debates. Finally, the strategy aims to support the management of the project, supporting good relations and building trust between key stakeholders – especially around the methodological rigour and independence of the evaluation.

Figure 15: Process of developing the Communication Strategy



6.1 Communication goals

238. There are three key goals to the communication strategy, each of which relates to a different target audience. These are:

1. Policymakers (in multilaterals, bilaterals and governments), the 'evaluation sector/community' and the interested public (via the media and appropriate fora) are aware of the Impact Evaluation and have a grounded understanding of its progress, results and policy implications.
2. Researchers' engagement with the Impact Evaluation's methodology and results affirms its rigour and enables application of these approaches to other contexts.
3. Project partners and stakeholders are aware of project progress, results and implications (and are able to use this in their own communication and reputation management strategies).

[42] See for example, Clemens, M. and Demombynes (2010), op cit. A summary of blog debates in 2001 can be found at: http://www.aviewfromthecave.com/2011/10/empire-strikes-back-sachs-vs-world.html.

6.2 Communication objectives

239. This Chapter expands the communication objectives for each of the goals outlined above.

Goal 1: Policymakers (in multilaterals, bilaterals and governments), the evaluation sector and the interested public are aware of the Impact Evaluation and have a grounded understanding of its results and implications

240. Policymakers and the Evaluation sector:

- Map out **target audiences** (identifying specific people; or positions or key institutions if we don't know the detail), our degrees of separation to them, and use outcome mapping principles and undertake **stakeholder analysis** to design influence strategy, and overlay this over communication timeline (see Table 17, below).

Examples of 'policymakers' in this context include: *national and local government* (including relevant ministries such as health, education, agriculture), SADA, and the district authorities; and, *donors* such as the Department of Foreign International Development in Ghana, World Bank, UNICEF, DFID, etc.

Examples of 'the evaluation sector' include *individuals* (such as Heads of Evaluation in NORAD, Danida, etc.); *networks* (E.g. XCeval Yahoo! Group – an active online discussion group which shares sector information and provides peer support); and, *organisations* (such as 3ie, and the African Evaluation Society).

- Map out and develop a **timeline for influencing opportunities**
- Identify appropriate **communication channels** for reaching target groups both *in shared spaces* and in *created spaces* to promote the MV evaluation. Examples of communication channels include: annual national workshops in Ghana (inviting key stakeholders); local and national media, online portals and knowledge sharing services (e.g. Eldis, R4D).
- Prepare and make available/disseminate **suitable communication tools** (e.g. policy briefings, project website, FAQs) providing clear information on the project and highlighting policy implications, when relevant,[43] and cross reference with timeline to generate/respond to **opportunities for appropriate communication** of progress and results.

Table 17: Example of timeline under development (for early 2012)

	March	April	May	June
External pegs			• Official launch of Ghana MVP	• Rio+20 ; G20 Summit
Project landmarks	• Draft Comms Strategy and Comms Plan	• Comms Strategy agreed • Develop reputational risk management strategy	• Project launch • Baseline survey in MV sites and control sites	
Media work		• EAG agrees messages • Draft media plan and identify media contacts (particular effort to engage with West African media)	• Project media briefing around project launch • Prepare FAQ for media queries around MVP's media release	
E-comms		• Investigate options project web presence		

[43] This may also require additional costs for translation.

241. The interested public:[44]

- **Identify 'warm' media contacts**, and media spaces which are likely to be receptive to briefings (from existing IDS and Ghanaian partners' media contacts, with input from EAG)
- Develop a 'light-touch' **presence in public spaces** and amongst NGOs with an interest in MV (e.g. the MV Facebook group)
- **Agree messages**, prepare media briefings and develop media contacts lists, assign project 'spokes-people' for media interviews
- Identify and respond to **media opportunities** (see external pegs on Communications Plan timeline), additionally, monitor wider media context (e.g. through IDS daily Media Briefing process)
- Develop a **reputational risk management** strategy for handling 'bad press', putting right incorrect messaging, address any media training needs that project spokesperson/people may have
- Ensure **project website** includes pages for a non-specialist audience, has good search engine optimisation (e.g. using keywords) so can be easily found via Google search, and is promoted through all project communication activities and through partner communication channels

Goal 2: Researchers' and the evaluation sector's engagement with the Impact Evaluation's methodology and results affirms its rigour and enables application of these to other contexts

242. Researchers (including consultants, think tanks as well as academics) and the Evaluation Sector:

- **Invite comment and debate** from other experts (e.g. invitation to be part of annual review panel, via project blog, invitations to participate in annual Ghanaian events, through commissioned articles for IDS Bulletin, contacting key bloggers and having them 'on board' the project) – some mapping may be involved first. Examples in Ghana include those from the research community, such as the Institute of Statistical, Social and Economic Research (ISSER), the University of Ghana, and Innovations for Poverty Action (IPA), as well as others like the Ghana Statistical Service (GSS). At an international level it may include: the UN Evaluation Group (UNEG), the International Development Evaluation Association (IDEAS), the World Bank, the Network of Networks Impact Evaluation Initiative (NONIE), and others.
- Identify **target spaces,** such as peer-reviewed journals (to submit articles), relevant seminars/conferences at which to present interim and final results and shared spaces for debate, and also make the most of existing **communication channels** (e.g. IDS Yellow Monday and E-Alerts).
- Prepare and make available/disseminate **suitable communication tools** (e.g. working papers, annual reports, the IDS Bulletin, presentations at relevant academic conferences) outlining methods and findings and cross reference with timeline to generate/respond to **opportunities for appropriate communication** of progress and results.
- Ensure **project website** contains project data and up to date information on methodological approaches and results (and where appropriate, to share these).

Goal 3: Project partners and stakeholders are aware of project progress, results and implications

- Ensure appropriate sign-off systems are in place for sensitive communication outputs
- Maintain regular internal communication amongst project partners and stakeholders (e.g. the EAG), engage communication point people (to be agreed) to flag up when new information is available or when new communication tool has been developed (with supportive materials like FAQs)

[44] These are primarily reached via the media and targeted spaces, like Millennium Villages website and communities.

6.3 Next steps

243. There are some important next steps as part of turning the Communication Strategy into a plan of action. These are covered under the following sections around agreeing appropriate messages, managing the communication risks, establishing an M&E system, and finalising the Communication Strategy.

Agreeing appropriate messages

244. There are a number of communication messages that need to be agreed for the MV evaluation. The wording of these will become an important element for ensuring consistency and clarity around the evaluation for different target audiences. This will need to include 'appropriate messages' around:

- **The purpose of the Millennium Villages** – both broadly and specifically in the context of Ghana, and background on why communities have been chosen to be part of the MV project
- **Impact Evaluations** – what they are, how they work, how they differ from evaluations, how this impact evaluation ensures independence and rigour
- **DFID's role as funder of the MV Impact Evaluation** – linking the above two: the importance of the MV project, and the value of rigorous IE with an emphasis on meaningful results for participants and value for money
- **Branding –** (how to describe who is doing what), and when to use/not to use logos

Communication Risks

245. There are a number of risks around this evaluation, which are both risks specific to the communication strategy (such as around wrong messaging), as well as risks that different communication methods may help mitigate. Indeed, there is a potential for criticism around the following areas:

- MV being example of Western interventionism/neo-colonialism, why certain communities were chosen over others, why Ghana again
- "Aid dependency"– how aid does not allow for local enterprise and business to build its resilience, creating dependency and being unsustainable
- Addressing the long-standing academic debates questioning the value and the approach of MVs (Does MV raise people out of the 'poverty trap' in the long term? Is MV more cost-effective than alternative interventions? Are there enough spill-over benefits to justify the significant investment on relatively few? Etc.)
- IE being seen as a 'waste of tax-payers money' – if we're not sure whether intervention works, why are we spending money on it; the IE is conflated with the MV itself and does not have sufficient independence

246. Clear messages by all partners should mitigate some of this, as will use of common tools such as FAQs, as well as ensuring transparency and accessible (in terms of language) communication. In addition, the reputational risk management strategy will clarify roles and responsibilities, assign communications point people and identify training needs.

Monitoring and evaluating the Communication Strategy

247. A range of systems exist (or need to be set up), to monitor quantitatively (e.g. publication downloads, web number visitors, number of media mentions, citations in academic journals) and qualitatively (e.g. feedback at annual conferences via evaluation forms, media monitoring analysis, comments on blogs, RTs and #FFs on Twitter, etc.). To this end, we will use a combination of Google Analytics to analyse the website hits (both for downloadable documents and comments on the blogs), as well as drawing on IDS' routine work in tracking the media and the messages it produces.

Agreeing protocols with the EAG

248. There are a number of elements around communication which will be agreed with the EAG over the coming months. These include:

- The role and level of engagement of EAG on communications, such as:
    - Who and how to sign-off sensitive messages (with appropriate support from the DFID-Ghana and the Press Office)
    - How to establish 'point communication people' (amongst key stakeholders)

- Finalisation of the Communication strategy, including:
    - Communications goals: anything missing? Do these need to be tweaked?
    - Communication objectives: anything to add? Any concerns that have not been addressed?
    - Messages, including agreeing the 'wording' and branding (these will be included in the FAQs)

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 get rid of 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 5 years. The MV model is already being implemented in a range of sites across sub-Saharan Africa, where it is now entering a second 5-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[45], leading to criticism of the results attributed to the MVs and the lack of independent rigorous evaluation[46]. 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 MV. The evaluation will cover up to a 10-year period – subject to programme renewal, – to answer evaluation

[45] Millennium Promise (2010), "Harvests of Development in Rural Africa: The Millennium Villages After Three Years".

[46] 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.

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 5-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 spill-over 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 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*?[47]

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

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

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-Difference approach with mixed methods is the most likely approach to be able to meet the criteria.[48]

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.[49]

4.6. Surveys at the MV site and comparison sites will take place in year 1 (baseline) and at least twice more during the 5-year MV direct implementation period.

4.7. Spill-over 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 3. This will determine if there is sufficient evidence of progress against its objectives to justify completing the full funding to year 5. 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.
- 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

[48] 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.
[49] The field report contains the current village matching checklist used by the MVP

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.[50] 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 later years as evidence is emerging. 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 5 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 5 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 M&E for the northern Ghana site, and associated MV organisations[51]. 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 5-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.

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

[51] 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.

6.4. We do not want to be overly prescriptive on staffing arrangements but expect the evaluation team would 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 hour's 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 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.

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 establishing 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 5-year period.

8.2. The 5-year MV intervention is scheduled to start its set-up phase in late 2011, with a detailed design phase of up to 6 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 establishing 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 5 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 5 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:

- An initial design document within the first 6 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.

- Baseline surveys completed within the first 6 months of the implementation of the MV project

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

- 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.

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

- '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.

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

10. DFID co-ordination

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[52]. 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[53]. There have been concerted efforts for decades to reduce the stubbornly high rates of poverty in the North[54], but with little success[55]. The region exhibits the characteristics of what Jeffrey Sachs calls a 'poverty trap' deriving from a paucity of various forms of capital[56]. The Government of Ghana acknowledges the particular challenges faced by the North, and in 2010 created the semi-autonomous Savannah Accelerated Development Authority (SADA)[57]. 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"[58].

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 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.

[52] Ghana Living Standard Surveys (GLSS) 3 and 5, conducted in 1992 and 2006.
[53] World Bank staff calculations, based on GLSS 5 in 2006.
[54] CEPA and ODI (October 2005), "Economic Growth in Northern Ghana", for DFID.
[55] World Bank (March 2010), "Tackling Poverty in the Northern Ghana".
[56] Jeffrey Sachs (2005), "The End of Poverty: Economic Policies For Our Time".
[57] Government of" Ghana (2010), SADA Act Number 805.
[58] SADA (2010), "SADA Strategy and Work Plan 2010 - 2013: A Sustainable Development Initiative for the Northern Savannah", Savannah Accelerated Development Authority.
58 www.millenniumvillages.org

11.3. The first MVs commenced in 2006[59]. 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[60]. 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[61]. 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 the North of Ghana. It fits well with DFID's increased emphasis on innovation and on achieving real development results[62]. And it would be in line with DFID-Ghana's new Operational Plan that proposes increased focus on the poor North of the country. 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 the MV site in northern Ghana. However, given the innovative nature of the approach, and the high-profile debate that has surrounded it[63], 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.

[59] 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.
[60] Millennium Promise (2010), "Harvests of Development in Rural Africa: The Millennium Villages After Three Years".
[61] 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.
[62] See for instance, DFID (2011), "UK Aid: Changing lives, delivering results".
[63] 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/).

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[64]. 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:

1. longitudinal household-level assessments over time
2. periodic assessment of impacts against interventions
3. non-randomized 'plausibility' evaluation against a separate local matched comparison group
4. comparison against a separate intervention such as cash transfers
5. comparison against regional trends
6. "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 which is comparable to that of rural Ghana, including field work. 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

[64] 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".

14.1. Bid responses should not exceed 50 pages (size 12, single line-spacing), 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. Theories of change and initial evaluation questions

The impact evaluation team has developed Theories of Change diagrams for the sectors of health, agriculture and education, based on the impact pathways set out by the MVP and the MV Handbook. The logic underpinning the sectoral interventions is complex, due to the multiplicity of types of activities implemented under each sectoral intervention area, as well as the complex inter-relationships and synergies between the sectors.

For each sector the associated theory of change sets out the intervention logic and illustrates the vertical and horizontal inter-relationships.

Vertically the diagrams illustrate the hierarchy from outputs, outcomes to impacts, which help break down the causal linkages between these different levels in the hierarchy. While the impact level focuses on the MDGs, the outcomes level focuses on the short-term and intermediate outcomes and illustrates in an incremental manner how the outcomes work towards the impact level in the intervention logic. The assumptions underpinning these causal linkages are mapped into the process, as essential preconditions driving the generation of the impacts expected from the interventions. A key assumption underpinning the three sectors (agriculture, health and education) that have been examined is the receptiveness of the community or the specific beneficiary group to new ideas – which should drive consequential changes in knowledge, attitudes and behaviour. The MVP is driving forwards a number of changes in practice (from changes in family planning practices and nutrition, to changes in crops produced and farming activities, to changing attitudes towards gender). Testing the validity of these underlying assumptions will be important part of the qualitative assessment conducted in the framework of the impact evaluation.

Horizontally the inter-relationship between the interventions is illustrated and typically shows an increasing degree of *systemic change* as the diagram moves from left to right. For example the interventions in the agriculture sector move from 'quick win' -type interventions (e.g. the distribution of fertiliser and seed crops) towards agricultural diversification and improvements in market infrastructure and systems. Similarly in education, the programme moves from increasing primary school enrolment through to reducing barriers to secondary education and improving the skills of those not presently in formal education.

Crosscutting themes feature very strongly in the intervention logic for the different sectors. In particular community engagement and empowerment is central to the MVP approach to achieving the MDGs. This community-based approach stems from the principles of participation, social and gender inclusion, equity and local stakeholders' ownership of the decision-making and development process. Thus the community engagement and empowerment is a key driver of the sectoral theories of change, and will feature prominently in the qualitative assessment of impact as well as likely sustainability of the MV project.

Agriculture ToC

Eradicating extreme poverty & hunger (MDG 1)

Ensuring environmental sustainability (MDG 7)

Improved food security

Increased income

Markets for agricultural outputs are improved

More healthy population

Production of food is increased and sustained

A1

A3

A6

Increased access to financial services and products

Environmental impacts of farming systems reduced

Increased capacity of farmers & comm instits to manage benefit from ag activities

Improved agricultural production risk management

Improved water management (link to water pathway)

Increased use of improved staple food crop production practices

Increased access to improved seeds & fertilisers from private agro-dealers

Financing/credit systems for distributing ag inputs

Info campaigns & supplier fairs for agro-dealers

Agro-dealers trained on accessing improved inputs

Micro dams & ponds installed

Incentives for households to adopt micro-irrigation

Drip irrigation systems installed

Extension agents trained

Comm structures & ex agents trained on micro-irrigation maintenance

Comm structures & extension agents trained on increasing water availability

Comm structures & ex agents trained on controlling ag pests and diseases

Drought and pest-resistant crops promoted & introduced

Ex agents trained on interpreting & communicating met forecasts & making recoms on crop management

Farm demonstration sites

System to improve women's access to agri inputs

Voucher system established

Farmer field Visits facilitated

Farmer groups & extension agents trained on small-farm management practices

Farmers & ex workers trained in improved soil mgmt

Nat & I'nat micro-finance insts identified

Market and finance Service provision studies facilitated

Comms structures Tained on financial mgmt

Cap development for ex agents & comms orgs to coordinate ag activities

Meetings for women farmers to understand their problems

Promoting participation of Women in ag planning

Safety nets are improved for crop failures, post-harvest and livestock losses

A4

Increased reserves of staple crops

Enhanced food risk and crisis management

Enhanced access to drought insurance scheme

Comms based orgs trained on risk mgmt

Farmers & ex agents trained on storage methods

Cereal banks constructed

Rainfall stations established

Drought insurance index plan submitted for financing

Training for comms Orgs in interpreting Climate data

Training & tech Support to comms to manage insurance scheme

Increased access to livestock re-stocking schemes

Households requiring re-stocking identified

Linkages with gov or NGO re-stocking schemes facilitated

Income generation from agricultural activities is increased

A2

A4

A5

Increased number of buyers and sellers

Increased no. of small-scale processing, handling and other value enterprises

Increased sales of staple foods

Increased production and productivity of high value agricultural commodities

Market studies for new high value crops

Demonstration plots for high-value commodities

Comms orgs & ex Agents trained on tech & mgmt of high -value commodities

Incentives for Adoption of new commodities

Access, allocation & role of women in decision making with regard to ag income

A5

A6

Increased cap of farmer orgs to get market info,negotiate contracts & meet demand

Improved market infrastructure & info for ag products

Cap dev for households to manage income from ag activities

Incentives for adopting new value adding enterprises

Financial service provision studies of potential service providers facilitated & sponsored

Entrepreneurs trained On tech & mgmt of small scale processing

Interactions between farmers & traders facilitated

Comms-based orgs trained on marketing of staple foods

Market studies for staple foods

School-feeding programs explored

Contract farming (or similar) developed

Farmers trained on contract design & law

Farmers linked to potential buyers of produce

Major roads linking cluster with major markets repaired

Transport of ag products to markets facilitated

Market trading centres repaired &/or built

System to collect & analyse market info

Market studies for new value adding enterprises

Consumption of balanced Diets Is increased and sustained

A6

Increased adherence to exclusive breastfeeding & use of clean water & bottles (see health pathway)

Improved and diversified dietary intake

Health staff, ag ex agents & comms trained on nutritional issues

Info on diet provided to health facilities, schools & comms centres

More nutritious & diverse crops promoted & introduced

Health facilities & personnel provided info on breastfeeding

Water treatment for household drinking facilitated

Medium-term interventions

"Quick wins"

Staple crop production

Risk management & coping strategies

Diversification

Market infrastructure & access

Diet & health

Millennium Village Hunger, Agriculture & Nutrition Sector
Theory of Change Diagram Legend

	Outputs
	Intermediate outcomes
	Ultimate outcomes
	Impacts
	Intervention groupings
	Assumptions

Note on the sector's intervention logic

The intervention logic for agriculture and nutrition illustrated in this theory of change diagram incorporates vertical and horizontal inter-relationships.

Vertically, a relationship exists from bottom to top in the diagram, which progresses from outputs at the lower end of the intervention hierarchy, through two levels of outcomes, to impacts at the top. This relationship connects the planned activities in this sector to the expected results at the outcome and impact levels, including results relevant to MDGs 1 & 7. The outputs are also sub-divided vertically into “quick wins” and medium-term interventions to reflect the prioritisation of activities during the 5 year implementation period. It is understood that “quick wins” will encompass early activities which, to a large extent, will be implemented in the first year before the first harvest.

Horizontally, a relationship exists from left to right in the diagram, with interventions to the right aiming for an increasingly degree of systemic change beyond existing agricultural activities. Interventions have been grouped into a number of thematic areas. Those on the left are focussed more on improving existing farming activities and in managing risk associated with these. As a result, the majority of the early “quick wins” are located to the left of the diagram. Those activities to the right of the diagram aim for a transition away from traditional farming relationships based on staple crops towards commodity crops, increased market integration, and diversification within and without farming. The main exception to this typology is the final grouping, “diet and health”, which contains elements of other groupings but is largely stand-alone, and links the agriculture sector to the health sector.

Agriculture assumptions

A1	That farmers and communities are open to new ideas and ways of working and living.
A2	That increased income gained from increased production, greater market access etc wil translate into reinvestment to sustain income increases into the next years
A3	That the training, incentives, structures and systems created will be used effectively and taken up by the host communities in sufficient volume to realise change
A4	That the interventions will empower the communities to participate and run groups, projects, businesses be involved in decision-making etc. both during and after the intervention
A5	That the new crops, other agricultural activities, ways of working promoted by MVP are appropriate to the local context, farming systems, cultural norms etc.
A6	That other partners and stakeholders will co-operate with MVP in the delivery of the programme such as micro-finance institutions, schools, health facilities.

Example of detailed Evaluation Questions (agriculture):

Outcomes EQs		Process EQs	
Increasing agricultural yields and diversifying crops are critical steps in the MV approach **Income/Production Change** • What is the change in income from agricultural activities? • By how much has production by farmers of food by crop type increased? • What is the change in the number and associated yield of new and additional crops? • What is the level of change in the production of high value agricultural commodities? • What is the number of high value agricultural commodities produced? • What is the change in the ownership of livestock by farmers? • What is the level of change in the volume of sales of staple foods? • How many small scale enterprises who are processing or adding value to the agricultural outputs are participating in the MVP? HH = Household Questionnaire M = Male Questionnaire F = Female Questionnaire	**Data Source** **Baseline Questionnaire** **HH Module S2** **HH Module S2** **HH Module S2** **HH Module S2** **HH Module S2** **HH Module T** **HH Module S2** **HH Module P** **(NB: the**	**Timing and Sequence:** • Was the timing and sequence of the interventions appropriate? Was the most efficient and effective approach to planning the roll out of the interventions taken? • Were the right interventions prioritised? • How were the interventions customised to the local context (agro-climatic zones, farming system, cultural context)? • Is the package of services provided by the project (supply of seeds, fertiliser, training, establishment of village farmer organisations) *sufficient* to generate the results expected in terms of increasing agricultural productivity and diversification. **Uptake** • What is the proportion of farmers using improved seed and fertilizer (MVIS performance indicator)? • How many farmers have benefitted from the livestock interventions? How were these farmers selected? • What are the barriers/drivers of uptake of the interventions by the community, farmers, HH etc. • How did the MVP select beneficiaries? How were women targeted? Were vulnerable groups targeted? • Did the MV interventions lead to more market access for farmers or for small businesses? • Are particular groups of farmer / profile of farmer gaining more than others? How is equitable access ensured? **Training and Capacity Building** How effective are the training /demonstration plots provided in promoting uptake and ongoing use of the ideas /systems promoted? How effective have the training interventions been in developing sustainable capacity to increase agricultural productivity and support agricultural	**Qualitative Data collection** Was the timing and sequence of the interventions appropriate? Were the right interventions prioritised? How were the interventions customised to the local context? Is the package of interventions sufficient to generate agricultural diversification and increased agricultural productivity? What are the barriers/drivers of uptake of the interventions by the community, farmers, HH etc. How did the MVP select beneficiaries? How were women targeted? Were vulnerable groups targeted? How effective have the training interventions been in developing sustainable capacity to increase

	question in questionnaire is more focused on non-ag enterprises)	diversification? How were the training participants selected? How were the training topics selected? Was the training adapted to the local context and training needs of farmers? **Market Dynamics** • What have been the improvements in input and output markets for agricultural products? • How is the system of input subsidies/vouchers working in practice? • Is the project facilitating increased access to fertilisers and seed varieties by agro-dealers? Are local agro dealers able to tackle the volumes of materials required and have the distribution channels needed to reach the target population? Is the project succeeding in driving changes re local input markets? • How is the dynamics of supply and demand playing out in the markets for new crops? (price, livelihoods, incentives for farmers to continue)? • What are the key drivers of improved market access (roads, transport, market centres, market information)? Are these being addressed by the project? • Are farmers using the storage facilities to store crops post harvest in order to secure better prices later? • How effective has the support to community-based organisations in the marketing of staple foods? • Has the MV package of interventions caused the conflict between farmers and people raising cattle to decrease? • What is the level of value added by the small scale enterprises? • How is the demand to supply schools dovetailing with supply side issues? **Agriculture Committee/Farmers organisations/Co-ops** • How many agricultural committees are there in the MVP cluster? How many people participate in these committees (according to gender, etc.) • What is the configuration of the membership of the Agriculture Committee? How are members selected? Is there any turnover of members? Local political involvement? Gender balance? • Is the training and support provided by MVP to the Committee and other ag groups good quality, sufficient?	agricultural productivity and support agricultural diversification? What have been the improvements in input and output markets for agricultural products? How is the dynamics of supply and demand playing out in the markets for new crops? (price, livelihoods, incentives for farmers to continue)? What are the key drivers of improved market access? How effective are the community based committees/co-ops / associations in securing local ownership of MVP agricultural

		• How many *functioning* agricultural co-operatives have been created? When were they established? What is their purpose (marketing of crops, adding value through processing, etc.) • What is the membership of these organisations (by gender, socio-economic group, etc.)? How often do they meet? Where? • How effective are the co-operatives involved in agriculture or broader business development? • Is the relationship between the community based associations and MVP changing as the former get more embedded and gain more experience? • How active are the community led organisations in defining priorities and in making and implementing decisions? • Are the systems being created (such as the selling of produce to more distant markets supported by community organisations) working? • Are the structures for the co-ordination of agricultural activities working well? (e.g. Agriculture Committee, Grain Bank Committee). How could they be improved? Lessons? • What are the drivers of community participation in these structures? Are these structures/groupings sustainable? Why?	interventions and input into MVP decision making and prioritisation processes? What is the value added by the community based associations/organisations? What are the drivers of community participation in these structures? Are these structures/groupings sustainable? Why?
Achieving calorific food security is a first year goal of most MVs via increasing agricultural productivity initially; later via diversification. **Food Security** • What is the change in the proportion of food insecure HH[65]? • What is the change in the level of malnutrition (more specifically % of under fives who are underweight) in the MV?	HH Module G	• What is the change in the volume of reserves of staple crops per farmer? • What is the volume of surplus stored in the Cereal Banks/grain storage structures? • How many HH use the Cereal Banks, HH and community grain storage structures? • What is the change in the acreage of planting of pest and drought resistant crops? • How many farmers are participating in the drought insurance scheme? • Who are the most food insecure or vulnerable population groups? • Where do HH obtain their food? • What are the factors which limit their ability to obtain food from these sources? • How has the package of MV interventions generated more food security for the people of the MV?	Does the MVP approach do enough to address food insecurity? What improvements could it make to address this issue, looking specifically at particular groups in the community. Have the early agricultural interventions worked in tackling food insecurity in the early years? What interventions have proven

[65] MVP Handbook suggest this can be calculated by assuming 100kg per person per year of a staple cereal crop, multiplied by the number of people in the HH.

		• Is the process of addressing the unmet needs of food insecure HH post-harvest working? How and why? • Have the early agricultural interventions worked in tackling food insecurity in the early years? What interventions have proven to be most effective and why? • Does the MVP approach do enough to address food insecurity? What improvements could it make to address this issue, looking specifically at particular groups in the community.	to be most effective and why? How has the package of MV interventions generated more food security for the people of the MV?
Environmental Sustainability • How many farmers have improved access to water for agricultural activities? • What is the number of farmers practicing improved soil management techniques?	HH Module S5 HH Module S3	• What is the proportion of farmers benefitting from small scale irrigation programmes? (MVIS performance indicator) • Are the small scale irrigation systems proving effective in improving the availability of water to farmers? • Are the effects of flooding reduced and how? • Is the water supply more reliable throughout the year? • Is there sufficient access to equipment required for ongoing water management such as rainwater harvesting and storage, low pressure irrigation systems, etc. • What is the number and capacity of the rainfall stations established? • Who is responsible for maintenance of systems and ongoing water management? Who will fund these in the future?	Are the effects of flooding reduced and how? Is the water supply more reliable throughout the year? Who is responsible for maintenance of systems and ongoing water management? Who will fund these in the future?
Improving nutrition and nutritional diversity **Health** • What is the proportion of the population with a diverse diet? • What is the level of change in adherence to exclusive breastfeeding? • How many schools are participating in the school feeding programme? • What is the number of children benefitting? • Numbers of HH benefitting from water treatment of drinking water?	F Module G F Module F HH Module C HH Module H	• How have attitudes and behaviours towards diet and nutrition changed? • Which methods are proving successful in the promotion of improved and diversified dietary intake? • How has crop diversification worked to support nutritional security? • Are particular groups "better" at putting the dietary advice into action? What are the factors which contribute to success? • What are the factors driving change in relation to the exclusive breastfeeding?	How have attitudes and behaviours towards diet and nutrition changed? How has crop diversification worked to support nutritional security? Are particular groups "better" at putting the dietary advice into action? What are the factors which contribute to success?

Example of cross-cutting process issues (from agriculture theory of change)

Synergies	• How are the interdependencies between the different outcome areas working in practice? Cross sectoral linkages concerning the agricultural sector are listed below: - School meals programmes (increasing attendance and performance in schools) - Health; - Integration of soil fertility management practices to avoid contamination of surface waters - Environment; - Small scale irrigation techniques supporting high value crops and ensuring better HH nutrition - Health; - Increased participation of women in agricultural income management to ensure increased income is invested in the HH – Gender, Health; - Cereal banks, input stores and financial services to develop markets and enable transition to small scale entrepreneurship (from sub-subsistence) - Business; - Protein rich diets for HIV/AIDS patients to ensure effectiveness of antiretroviral drugs - Health. • Could the linkages between certain interventions be strengthened and what value would this add in terms of impacts? • Are there any trade-offs or negative repercussions emerging from this integrated approach? How can these be prevented/minimised? • Are the MV interventions joined up with what the Government/NGOs are doing – for example concerning restocking or other areas?	How are the interdependencies between the different outcome areas working in practice? Could the linkages between certain interventions be strengthened and what value would this add in terms of impacts? Are there any trade-offs or negative repercussions emerging from this integrated approach?
Community involvement and local ownership	• How was the MV model adapted to the needs of local context and in what way is the local community involved? • Is the local community supportive of the ideas and approaches taken? What are the issues? Lessons learnt? • Is the MVP approach succeeding in creating a locally driven partnership, new structures and systems with strong buy in from local communities? • Are local schools, health facilities buying into the programme and actively participating? • How is local government engaged in the MVP e.g. setting of priorities, implementation? Is local government supportive of the MVP approach? Why? • How is the process of working in the two Districts impacting on the implementation of MVP and the	Is the MVP approach succeeding in creating a locally driven partnership, new structures and systems with strong buy in from local communities? How is local government engaged in the MVP?

	establishment of community systems and structures? Is there any difference in the allocation of resources across the communities in the two districts? • Is SADA involved in the project/engaged with MVP in the delivery process and examining results? • Are other neighbouring communities/districts (not in MV) showing interest in the MVP interventions? • How is the capacity of local government, key agencies and local institutions being enhanced? • What are the drivers behind local policy change and ongoing commitment to the processes started by MV?	How is the capacity of local government, key agencies and local institutions being enhanced? What are the drivers behind local policy change and ongoing commitment to the processes started by MV?
Gender issues and vulnerable groups	• How have the agricultural interventions benefitted women? Increased income, more diversified sources of income, better diet and nutrition? Have there been any negative impacts/unintended impacts? • How has the MVP approach secured the involvement of marginalised and vulnerable groups? How have they benefitted/been affected? • Which groups have not benefitted? Why? • What have been the issues in terms of engaging with women and other marginalized groups? Were these issues addressed appropriately and effectively? • How has MVP worked in terms of generating differential impacts on men and women? • Has MVP impacted on gender relations and the roles of men and women in the HH? • Have there been any barriers / inhibitors impacting on the participation of particular groups on the project?	How have the agricultural interventions benefitted women? How has the MVP approach secured the involvement of marginalised and vulnerable groups? How have they benefitted/been affected? How has MVP worked in terms of generating differential impacts on men and women? Has MVP impacted on gender relations and the roles of men and women in the HH?

(Note: The key assumptions are detailed as part of the Theory of Change diagram)

Health ToC 1: treating/decreasing cases of HIV/AIDS, TB & malaria, & other neglected tropical diseases

Combat HIV/AIDS, malaria, and other diseases (MDG 6)

Knowledge & awareness raising

Awareness & understanding of HIV/AIDS, malaria, TB & other tropical diseases transmission, treatment & prevention is enhanced

A2

Consistent & proper use of bed nets & reduction of mosquitoes in the household environment

Increased awareness of condom use for double protection (both FP & STD), knowledge of high risk sexual behaviour & sexual debu & reduce HIV stigma in the communities

Increased awareness & understanding of disease transmission, treatment & prevention incl when to seek care (all diseases

Health staff & VHWs trained on disease transmission, best practice treatment & prevention

Comms trained on disease transmission, treatment & awareness

Public awareness campaigns and VHW home visits conducted

Educational materials on disease prevention, control made available to the comm

Information made available in consistent & targeted manner (public campaigns, VHW visits etc.)

Public awareness campaigns on condom use for double protection conducted

Comms & health personnel trained on condom usage

Information made available in consistent & targeted manner (public campaigns, VHW visits, health staff, etc.)

Bed nets made available

Comms trained & sensitized on bed net use

Comms trained on reducing breeding sites

Homes sprayed

Monitoring of malaria levels & mosquito populations

Comms sensitized on NTD transmission areas & behaviours to avoid transmission

Diagnosis

Diagnosis of HIV/AIDS, malaria, TB & other neglected diseases & associated conditions is enhanced

A1

Increased availability of diagnosis of diseases & associated conditions (equipment, agents, trained staff)

Increased no. of people tested for HIV & eduated re primary & secondary prevention

Increased screening, availability & access to comms-based treatment of NTDs

Health staff trained on disease diagnosis

Equipment & reagents procured on an ongoing basis

Appropriate referrals made by VHWs for disease treatment

Increased no. of places within the community to get tested

Health staff trained on HIV/AIDS disease diagnosis

Diagnosis equipment & reagents procured on an ongoing basis

Stool sampling undertaken

Treatment

Availability of quality treatment of HIV/AIDS, malaria, TB, other neglected diseases & assoc conditions is ensured

A1

A3

Adherence to treatment is increased (all diseases) (e.g. ART, DOTS)

Efficient referral mechanisms to higher levels of care in place for all diseases

Increase compliance and cure rates for TB treatment

Increased demand, access & availability of quality treatment of diseases (malaria & TB) assoc conditions & PMTCT to reduce vertical transmission of HIV/AIDS (drugs, trained staff etc.) - clinical & therapeutic

Administration of the appropriate TB treatment

Disease-relevant drugs procured on an ongoing basis

Health staff trained on treatment & counselling

Appropriate protocols available & used

Referral protocols in place

Communication mechanism in place

Community health workers conduct outreach on drug adherence & support

CHWs trained on DOTS

DOTS supervised at home

TB/HIV patients identified

Millennium Village Health Sector Theory of Change Diagram Legend	
	Outputs
	Intermediate outcomes
	Ultimate outcomes
	Impacts
	Intervention groupings
	Assumptions

Note on the sector's intervention logic

The intervention logic for HIV/AIDS, TB, malaria and other neglected tropical diseases (NTDs) illustrated in this theory of change diagram is divided horizontally into three distinct intervention areas.

On the left of the diagram is knowledge and awareness raising, which encompasses training and education delivered to health workers and the community to improve awareness and understanding of disease transmission, treatment and prevention.

In the middle of the diagram is diagnosis which includes activities to enhance the knowledge of health staff of techniques for diagnosis, and the equipment and infrastructure to support this, with the aim of increasing the availability of diagnosis and volume of testing for HIV/AIDS and other diseases, and community-based treatment of NTDs.

To the right of the diagram is a group of interventions related to treatment which aim to improve the skills of health workers, availability of drugs, and adherence to international protocols to improve the quality of treatment and to the strength of the regional health referral system.

Health (ToC 1) assumptions

A1	Sufficient staff can be recruited for Village Health Worker roles and other health personnel roles.
A2	The community is receptive and open to increasing awareness raising activities and changing their attitudes and behaviours to the messages conveyed via the MVP programme (public awareness raising actions).
A3	Health sector stakeholders and partners (such as District hospitals) co-operate effectively with MVP to ensure joined up access to health care services.

Health ToC 2: enhancing nutrition and reducing health problems associated with nutritional deficiencies

Eradicate extreme poverty & hunger

Knowledge & awareness raising

- Health staff & VHWs trained on nutritious crops, supplementation campaigns, growth assessments & nutritional issues
- Public awareness campaigns on nutrition & VHW home visits & supplementation campaigns on critical Micronutrients conducted
- Edu materials on nutrition available in health facilities & at comm level
- Comms members trained on nutritious crops & nutritional issues
- Outreach nutrition edu campaigns for HIV/AIDS patients, OVCs & vulnerable groups in community conducted
- Improved dietary options promoted & distributed by schools, comms centers, health centers, & local leaders

Direct intervention to improve nutrition

- Water sources used for drinking in the household treated where necessary (water & san pathway)
- More nutritious & diverse crops promoted & introduced where appropriate (from agri pathway)
- Linkages between market traders & suppliers of nutrient rich food established
- Options for nutritious diets for school meal programs designed
- Strategies for procuring nutritious foods in school feeding programs designed

Ongoing diagnosis & treatment of nutritional deficiencies

- Health staff trained on nutritional deficiency diagnosis
- VHWs making referrals
- Needed diagnostic equip & regents procured on ongoing basis
- Mechanisms to procure & make nutritional supplements on ongoing basis set up
- Health staff trained on malnutrition treatment
- Treatment protocols for malnutrition made available in all health facilities
- Mechanisms to ensure vitamin A, iron/folate, iodine are available on ongoing basis for women & children
- Macronutient supplementation programs for vulnerable groups incl. widows, orphans, PLWHA developed in coop with comm
- VHW referral mechanisms for vulnerable groups developed & set up

A1 — Consumption of diets that meet min requirements for energy, protein, fat & essential vitamins & minerals is increased

- Knowledge of improved diets among comms
- Increase use of best practices on complementary feeding 6-24 months
- Reduced no. of malnourished children & increased growth of children

A2 — Consumption of diets that meet min requirements for energy, protein, fat & essential vitamins & minerals is increased

- Increased adherence to exclusive b'feeding or use of clean water & sanitised bottles
- Increase food and nutrition security at household level
- Increased no. of biodiverse crops & availability of nutritious foods on farms, in markets & households
- Schools provide nutritious meals to students

A3 — Consumption of diets that meet min requirements for energy, protein, fat & essential vitamins & minerals is increased

- Increased use of growth monitoring by schools, comm health workers & families

A1 — Diagnosis & treatment of nutritional deficiencies is improved

- Increased capacity for diagnosing nutritional deficiencies & assoc conditions

A2 — Diagnosis & treatment of nutritional deficiencies is improved

- Increased availability & access to quality treatments of nutritional deficiencies & assoc conditions
- Increased availability & access to nutritional & macronutrient supplementation, esp for pregnant women and children under 2

Millennium Village Health Sector **Theory of Change Diagram Legend**	
	Outputs
	Intermediate outcomes
	Ultimate outcomes
	Impacts
	Intervention groupings
	Assumptions

Note on the sector's intervention logic

The intervention logic for enhancing nutrition and reducing problems associated with nutritional deficiencies illustrated in this theory of change diagram is divided horizontally into three distinct intervention areas.

On the left of the diagram, the first cluster of activities is organised around raising the knowledge and awareness of health staff and the community of the benefits of, and ways of attaining, improved diets and best practices in supplementing children's diets.

In the middle of the diagram there are a number of direct interventions to improve food security and nutrition in the household and the availability of nutritious foods in farms and markets and schools.

To the right of the diagram, the final cluster of activities aim to enhance access to, and the quality of, diagnosis and treatment of nutritional deficiencies, as well as availability and access to macronutrient supplementation, by training health workers and enhancing the capacity and quality of health systems.

Health (ToC 2) assumptions

A1	Sufficient staff can be recruited for Village Health Worker roles and other health personnel roles.
A2	The community is receptive and open to increasing awareness raising activities and changing their attitudes and behaviours to the messages conveyed via the MVP programme (public awareness raising actions)..
A3	Co-operation with other partners such as schools, farmers etc. to ensure provision of more nutritious foods.

Health ToC 3: improved women's health, reducing maternal/child mortality

Reduce child mortality (MDG 4)

Improve maternal health (MDG 5)

A1

Improved maternal & child health incl reproductive health, reduced incidence of disease, illness & emergencies for expectant mothers

A2

Increased visits to ANC

Increased availability of & access to preventative malaria, tetanus & other prophylaxis & medical advice for pregnant women

Increased availability of diagnosis & treatment of maternal health

Improved household hygiene

A3

Increased awareness & understanding of maternal diseases, incl SRH

Health staff & VHWs trained on understanding of maternal & SRH

Comms members trained on maternal health incl SRH

Public awareness campaigns & VHW home visits are conducted

Educational materials on SRH & maternal & child health are available

Information provided to all households on hygiene issues

Training provided on hygiene, latrines, hand washing, etc (see W&S pathway)

Health staff incl VHWs trained on diagnosis of maternal & child issues incl need for EMOC

Necessary equipment & reagents procured on an ongoing basis

Referral mechanism functioning & VHWs referrals made

Staff member able to do EmOC surgeries trained & available in referral hospital

Referral hospital upgraded to perform EmOC services

Relevant drugs & medication procured on an ongoing basis, incl malaria, tetanus, & HIV

Health staff trained on treatment

Treatment protocols available

VHW home visits conducted to inform women about IPT & follow up usage

More ANC sites available that offer a full range of prenatal services

VHW home visits conducted to refer mothers to ANC

A1

Reduced mortality rates of mothers & babies & incidences of health risks during child delivery

A2

Increased institutional delivery/births attended by skilled attendants

Increased availability of skilled birth attendants with adequate equipment (midwifery trained nurses at village clinics)

Increased use of diagnostic tools like partographs & ability to diagnose for EMOC during childbirth

Information provided to all households on value of institutional delivery

Childbirth capacity at health facilities developed

VHW home visits conducted to inform expectant mothers

SBAs trained & available in cluster

Increased diagnosis & treatment of newborn health problems

SBAs, other health workers trained on diagnosing for EmOC during childbirth

VHWs conduct home visits & inform women about potential birth complications

Clean delivery kits available

SBAs & other health workers trained in the use of partographs

SBAs, VHWs & health workers trained in newborn care (warming, neonatal resuscitation, etc.)

A2

A1

Reduced infant mortality & improved health for infants

Increased exclusive breastfeeding or use of clean water & sanitized bottles for feeding

Increased availability of treatments for childhood diseases (oral rehydration therapy, diarrhea treatments, malaria, pneumonia)

Increased full immunization coverage for infants & children

Increased health care services for infants & children

A3

Health staff & VHWs trained on understanding of child birth

Comm members are trained in basic child health care issues

Educational materials on child health are produced & disseminated

Growth monitoring charts available to all mothers

Health staff trained on using growth monitoring charts

Mothers sensitized on importance of carrying growth monitoring charts for children

Immunization days held

Health facilities with vaccine fridges & vaccines in stock

ORT packets, anti-malarials, antibiotics available with VHWs &/or health facilities

VHWs trained on diarrhoea diagnosis & ORT provision & referral for pneumonia & malaria

Information on breastfeeding provided to all pregnant mothers

Training on clean water sources & sanitation techniques provided to all mothers who choose not to breastfeed

Reduced no. of births per woman through improved family planning services

A2

Increased demand for & uptake of family planning services incl early age child bearing & birth spacing

A3

Increased FP services & integration into pre & post natal services, incl availability of access to abortion services (where legal) & post abortion care (PAC)

Women & men increasingly participate in SRH decision-making, contraception, protection from STDs, GBV

Family planning services available in increased number of locations

Confidentiality of family planning services secured

Women informed about full range of family planning options, incl birth spacing

Awareness of the youth on marriage planning, family planning & contraception enhanced

Health staff trained on post abortion care services

Abortion (incl pre & post) equipment available in health facility

Health staff sensitized on gender

Women provided with information on their rights with regard to SRH

Men informed about family planning, contraception, protection against STDs & GBV

Maternal & child health

Child & maternal mortality

Family planning

Millennium Village Health Sector Theory of Change Diagram Legend	
	Outputs
	Intermediate outcomes
	Ultimate outcomes
	Impacts
	Intervention groupings
	Assumptions

Note on the sector's intervention logic

The intervention logic for improved women's health and reducing maternal/child mortality illustrated in this theory of change diagram is divided horizontally into three distinct intervention areas.

On the left of the diagram, a number of activities are focused on improving multiple aspects of maternal and child health, including: awareness of maternal diseases; availability of diagnosis and treatment for maternal health; increased availability of, and access to, preventative health care; and knowledge of improved household hygiene.

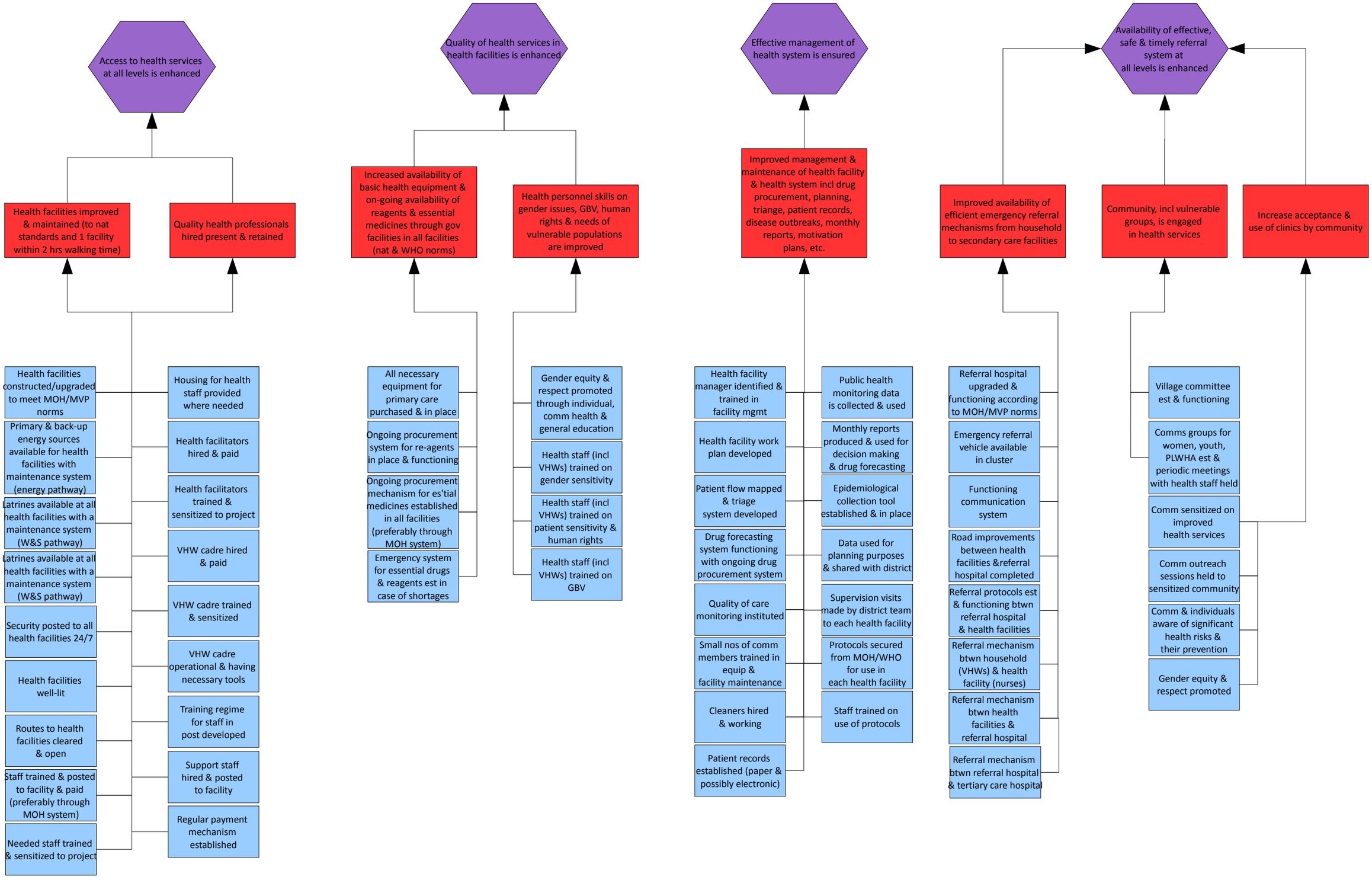
In the middle of the diagram there is a broad cluster of activities which aims to achieve a reduction in child and maternal mortality rates by improving diagnosis and treatment, improving health care standards for infants and children, increasing the number of deliveries attended by skilled attendants, achieving full immunization, and making treatments for childhood diseases more widely available.

To the right of the diagram a final cluster of activities is focused on family planning with the aim of reducing the number of births per woman by increasing the number of family planning services available and integrating them into pre and post natal services and encouraging women and men to increasingly participate in sexual and reproductive health decision-making.

Health (ToC 3) assumptions

A1	Sufficient staff can be recruited for Skilled Birth Attendant, Village Health Worker roles and other health personnel roles.
A2	The community is receptive and open to increasing awareness raising activities and changing their attitudes and behaviours to the messages conveyed via the MVP programme (public awareness raising actions).
A3	Health sector stakeholders and partners (such as District hospitals) co-operate effectively with MVP to ensure joined up access to health care services.

Health ToC 4: enhanced access to health services for all



Energy, Transport & Communications ToC

Reduce child mortality (MDG 4)

Improve maternal health (MDG 5)

Ensure environmental sustainability (MDG 7)

Eradicate extreme poverty & hunger (MDG 1)

Achieve universal primary education (MDG 2)

Promote gender equality & empower women (MDG 3)

Combat HIV/AIDS, malaria, and other diseases (MDG 6)

Develop a global partnership for development (MDG 8)

Access to clean energy sources for cooking

Availability of environmentally sustainable /neutral cooking & heating fuel sources is enhanced & sustained

Improved health from reduced emissions

Health risks associated with exposure to indoor air pollution (smoke) from cooking & heating are reduced

Efficiency of biomas use for cooking is increased (amount of fuel wood per unit of food cooked)

Availability of fuel wood near homesteads is increased

Comms structures & ex agents trained in establishing & managing woodlots & tree nurseries for sustained production of fuel wood, timber & fodder

Facilitating distribution of fuel wood, timber & fodder tress seedlings/seeds at comm & household levels

Cultural practices surrounding tree cutting & planting assessed (see environment & gender pathways)

Tree nurseries for fuel wood in comms & homesteads established

Support to comms-based orgs & ex agents on establishing & managing woodlots & trees for sustained production of fuel wood, timber & fodder

Demonstrations on the benefits of fuel efficient cook stoves

Training in "kitchen mgmt"/stove use provided

Linkages with (and training) groups engaged in efficient cook stove manufacturing, repair, marketing & use

Economic incentives for households & institutional kitchens (schools) to use improved stoves provided

Facilitating retailers to stock & sell kerosene & LPG in appropriate sizes/techs

Info on & demonstration of alternative energy sources & cost-effectiveness provided

Info provided & awareness raised on impact of smoke on health

Info provided on health benefits of using improved stoves

Costs of purchasing, installing & maintaining different electricity sources in comms assessed

Facilitating agreements for government contributions towards linking schools, clinics & comms centres to electricity grid

Assisting the acquisition of generators or solar systems in areas where the grid is too costly & for backup

Facilitating the installation of comms-scale electricity systems

Assisting in training comms-based structures & personnel in clinics & schools on maintaining generators & other electric equipment

Access to low-cost energy sources for community use

Enhanced access to improved & reliable low cost energy sources (for uses other than cooking)

Economic & physical access of merit (clinics, schools) & community centres & businesses to electricity is improved

Increased access to improved energy technologies enabling greater opportunities for income generation

Time and labour associated with mechanical power work is reduced (for grinding, milling, water pumping etc.)

Increased economic & physical access to improved energy sources for domestic communication, lighting needs

Economic & tech feasibility of improved energy techs assessed

Facilitating use of improved energy techs for income generation by hhs through training or new programs

Facilitating gov to assess, repair, build & improve all-weather roads network in the cluster

Working with comms-based structures to repair & submit requests to gov to repair & build main roads in the cluster

Materials & tools acquired where necessary

Spot improvement of roads, tracks & footpaths in the cluster

Comm mobilised & facilitated to carry out labour-based improvement & maintenance

Comms-based orgs trained on maintenance & repair

Facilitating repair of main road infrastructure for year-round access

Working with comms-based structures to prepare & submit formal requests to gov institutions for grid connection

Visits by gov officials to the clusters arranged & facilitated

Agreements for government contribution facilitated to link hhs to the grid

Economic & technological feasibility of different energy techs assessed

Subsidized energy techs acquired & distributed

Economic & tech feasibility of different home lighting techs assessed

Home lighting techs promoted

Areas where water pumping is economically /environmentally desirable identified

Comms-based structures trained on usage & maintenance of mechanical power systems

System to pay recurring costs of services established

Capacity development activities for comms-based structures & small-scale businesses providing grinding, milling, water pumping & agro-processing

Work with private sector providers to develop systems to support procurement

Community access to transportation links & services

Increased access to transport enabling continuous access to health facilities, schools, markets & public institutions

Improved transport infrastructure allowing minimum access to all-weather road network (<2km walking dist from hh)

Increased availability of affordable transport services to access key social & economic services

Facilitating provision of affordable & sustainable transportation services (for access to markets etc.)

Obstacles to accessing key cluster institutions identified & linking to key points outside

Comms-based, no-fee/low-cost emergency transport/ response system developed & implemented

Building capacity & raising awareness among comms-based structures on mgmt & ops of emergency transport services

Facilitating the development of a comms-based transport system for primary & secondary schools (comms > 10km away)

Capacity of comms-based structures on mgmt & operation of school transport services built

Management, operation & use of transportation systems & networks is inclusive & equitable

Representative committees for regulation & mgmt of comms transport services established

Comms-based structures trained on usage & maintenance of transport vehicles

Comms trained on the inclusive use of transport vehicles

Community access to information

Community capacity to access & disseminate information is developed through low-cost techs

Access to phones that provide public services within 2 km

Viability of mobile phone coverage assessed

Initial capital for purchase of mobile /comms phones provided

Expansion of cell phone network by providers s facilitated

Establish & support community radio station

Journalism & capacity building training provided

Active community radio station with community-wide access

Improved community access to outside sources of information

Viability of e-connectivity service assessed

Initial capital to establish e-connectivity provided

I.T. Infrastructure for comms learning, edu, health, business, & gov services established & supported

Working closely with other sectors to ensure sustainability & application-specific needs are met

Community sensitisation & training in ICT

Committee for regulation & mgmt for comms ICT services established

System for paying recurring costs of services established where necessary

Increasing proportion of people in emergency situations who are able to use emergency response system

Initial capital for emergency communications systems provided

Clinic officials trained on maintenance & mgmt of system

Comms sensitised/ trained in use of emergency system

Education ToC

Achieve primary universal education (MDG 2)

Promote gender equality & empower women (MDG 3)

Develop a global partnership for development (MDG 8)

Universal primary school enrolment achieved & attendance increased

A1

A2

A3

A5

Increased availability of quality teaching materials, aids, resources & supplies

Increased access to education by marginalised pops & students with disabilitiese

Children attend school tutoring & enrichment programs

Local barriers to primary education (esp for girls) reduced (see gender pathway)

After school tutoring programs established where needed

Quality teachers hired, present & retained

Improved skills on formal & informal teaching techniques

Direct & indirect costs of primary edu are reduced

Teacher performance & student achievement assessed regularly & reliably

Absenteeism due to preventable diseases reduced

Awareness campaigns on importance of primary & secondary edu

Quality teaching aids for teachers procured

Textbooks & learning materials procured

Comms members with skills to be mentors identified

Gaps to be filled & curriculum identified with teachers

Possible tutoring locations identified with community

School infrastructure improved & maintained

Incentives for teachers accepting rural postings

Support to comms orgs lobbying gov for minimum conditions for rural postings

Addressing gender imbalances in teachers

Comms early childhood edu initiatives supported & promoted

Scholarship Programs based on need &/or merit (gender pathway)

Teachers recruited & incentives to accept rural postings

Skill base of formally & informally trained teachers evaluated

Teachers trained in new & interactive methods

Teachers trained in new & interactive methods & content

Teachers trained in new & interactive methods & content

Teachers trained in Inclusion, gender & HIV/AIDS

Teacher support networks, exchange, linkages & workshops facilitated

Training sessions on nat level interventions created with MoE & local training colleges

Teachers trained on gender-inclusive pedagogy

Teachers use innovative eval techniques to monitor student achievement

Activities to monitor enrolment, attendance & student achievement

Activities to monitor Teacher skills & attendance

Activities to monitor Teacher skills & attendance

Classrooms built and/or rehabilitated

Pitlatrines constructed

Access to water

Access to power supply

Schools furnished

Single-sex pit latrines constructed

Boarding strategies for female children identified

School-based mass treatment (see health pathway)

Clean water, sanitation & hygiene awareness provided (see health pathway)

Working with local gov & donors to provide financial support to fill gap due to school fees

Creating strategies with other sectors to provide conditional transfers based on enrolment & retention

School feeding programs

Awareness campaigns on importance of educating vulnerable people by Comms Change Agents

Teachers trained in integrative & modicative strategies for children with special needs

Info on alternative schooling provided by NGOs & MoE for children w special needs

Incentives to students from marginalized pops & with disabilities

Flexible schedules & accelerated classes through alternative edu initiatives

Workshops for teachers & school admin in gender-based violence (gender pathway)

Enhancing school infrastructure & accessibility

Capacity of comms structures & education managers to monitor & evaluate formal & informal education activities strengthened

Community-based structures evaluate school mgmt, teachers & students

Parent-teacher associations & edu comms lobby for key activities to improve learning

School managers, PTAs & elected edu comms trained on M&E methods & tools

Training & tech support provided to school Managers & PTAs on fundraising, action plans, book keeping, Performance-based mgmt

Comms mentoring to monitor student attendance established

Governing charter to monitor school workings created

Evaluation of teaching & mentoring skills

Instruments to monitor post-training activities created

PTAs & Ecs conduct advocacy & awareness raising campaigns on self-defined issues

Policy dialogue between comms orgs & policy agencies/officials

Community engagement in education

Primary-school dropouts, 15-24 yr olds & child-bearing women are able to read, write, add & subtract (in official and/ or mother tongue)

A4

A5

A3

A6

Accelerated learning programs (ALP) created

Non-formal literacy & other education programs are established

Libraries established

Community actively supports non-formal edu for out of school children, youth & women of child-bearing age

Creation of clubs/ working groups with a literacy component integrated into the curriculum promoted

Youth & adult literacy classes in comms centrers offered

Work with other sectors to incorporate literacy training

Literacy curriculum for women's groups developed

Mentoring program for vulnerable youth provided

Holiday tuition summer enrichment programs

Vocational training programs for employment instituted

ALPs in the nat edu system established (in collaboration with MoE)

ALP curriculum integrated into vocational schools (in partnership with BDU sector)

Conditional transfers Identified & programs on attendance & enrolment for ALPs created

Education for out-of-school children

Transition from primary school to post-primary school edu is strengthened

Percentage of primary school leavers eligible for post-primary edu increases

Local barriers to non-formal post-primary edu are reduced

Local barriers to secondary edu are reduced

A7

A3

Awareness campaigns on the benefits of later marriages for girls conducted

Scholarship programs based on merit established

Assessments to identify vocational skills for trainings conducted

Meetings & campaigns for awareness of literacy programs conducted by comm

Edu committees work with marginalised comms to create programs which fit needs of pop

Mentors from the comm monitor progress of vulnerable youth (see Business/ Youth Pathway)

Library books procured through gov donations, partnerships with NGOs & other resources the comm can mobilize

Libraries housed in locations chosen by the comms

Teachers & mentors trained in integrating library resources into the classroom

Comm mobilised to organise, establish & monitor the library & renewal of resources

Scholarship programs based on need or merit established

School boarding strategies developed & implemented

Awareness campaigns on importance of secondary edu for girls conducted

Awareness campaigns on benefits of later marriages for girls conducted

Transition to secondary education

Millennium Village Education Sector Theory of Change Diagram Legend	
	Outputs
	Intermediate outcomes
	Ultimate outcomes
	Impacts
	Intervention groupings
	Assumptions

Note on the sector's intervention logic

The intervention logic for education illustrated in this theory of change diagram is divided horizontally into four intervention areas, which together encompass many different aspects of primary, secondary and tertiary/non-formal education.

To the far left of the diagram, the first cluster of activities aims to increase primary school enrolment by enhancing school infrastructure and the quality of teaching and by reducing barriers to access.

In the centre of the diagram are two groupings which aim to build, respectively, the capacity of school and community groups to manage and evaluate teaching outcomes, and to improve the skills of people who are not in formal education.

Finally, the group of activities on the far right of the diagram is to improve the transition from primary schooling to post-primary schooling by reducing barriers to secondary education and non-formal post-primary education, and by increasing the proportion of school leavers who are eligible for post-primary education.

Education assumptions

A1	Access to all types of schools is necessary to achieve universal primary enrolment
A2	That there are sufficient untapped resources to ensure that teacher shortages can be addressed eg others with at least some high school education can be trained to address teacher shortfalls
A3	Children and young people are willing to engage in the educational programmes envisaged for example mentoring, enrichment, life skills or problem solving programmes
A4	Local businesses, NGOs, microfinance institutions are willing to co-operate to provide training or employment opportunities
A5	Teachers are receptive to new teaching methods, programmes and meeting the needs of different vulnerable groups.
A6	The Ministry of Education is a willing partners, alongside other NGOs and partners.
A7	That the community is open to changing attitudes towards pre-existing gender norms – for example early marriage for females, role of education for females etc.

Water and Sanitation ToC

Combat HIV/AIDS, malaria, and other diseases (MDG 6)

Eradicate extreme poverty & hunger (MDG 1)

Reduce child mortality (MDG 4)

Promote gender equality & empower women (MDG 3)

Access to water

Increased access to water for households & public institutions

Improved reliability of supply from improved sources

Effort per person required to collect water is reduced (see gender pathway)

Increased supply of water available at public institutions

The supply of water for household consumption is increased & brought closer to homes

Participatory mapping of all existing water points

Participatory assessment of current water usage patterns, including cultural traditions & norms governing water collection & use

Improving water distribution capacity

Build and/or install new water storage facilities where needed

Train comms orgs & ex agents in techniques to reduce runoff & erosion & increase water infiltration, storage of water in the soil (see environment pathway)

Construct/develop terraces, trench bunds, etc. at the farm & landscape level (see environment pathway)

Facilitating construction/ development of new water points in accessible places for comms members

Protecting & rehabilitating existing water points

Train comms orgs on repair & maintenance of water point infrastructure & delivery

Repair & maintain existing water storage facilities

Efficient use of water

Increased efficiency of water use for agriculture (food production, cash crops & livestock)

Water supply for agricultural activities (including livestock) is increased

Quality of water for human consumption is improved

Treat water sources used for drinking, either on-site or at households (point-of-use disinfection)

Testing & monitoring for quality water at randomly selected households & sources

Distribute/facilitate acquisition of safe water storage vessels for households where needed

Treat on-site water sources used for drinking water

Test & monitor for water quality from key drinking points

Segregate livestock watering areas from water collections areas

Improvement s to physical water infrastructure to reduce contamination

Conduct awareness campaigns & training of comms-based orgs to reduce on-site contamination

Increased use of small-scale water management practices for agriculture, including livestock (with appropriate quality of water)

Train comms-based structures on efficient water management techniques & methods (see ag pathway)

Financial incentives facilitated to increase level of adoption of low-cost, micro-irrigation techs & small-scale irrigation techniques (see ag pathway)

Inputs for small-scale water mgmt facilitated/ distributed

Access to sanitation

Basic access to sanitation is enhanced

Increased availability of waste management facilities for households and public institutions

Increased use of sanitation facilities and hygienic practices

Increased use of appropriate waste management facilities

Increased availability of, and proximity to, an improved sanitary facility

Construction of waste mgmt infrastructure facilitated

Repair & maintenance of waste mgmt infrastructure facilitated

Linkages with broader waste mgmt systems & actors facilitated (see infrastructure pathway)

Cap dev activities for comms members on waste disposal & mgmt

Cap dev activities for comms-based orgs on repair & maintenance of waste mgmt facilities

Repair, improvement & maintenance of existing sanitary facilities for hh use facilitated

Construction of new ventilated improved pit latrines & other sanitary facilities facilitated

Construction or rehabilitation & maintenance of single -sex improved sanitation facilities at public institutions (see ed & gender pathways)

Participatory mapping of sanitary facilities, contaminated sites & sources conducted

Awareness/ed campaigns on hygiene health risks & benefits associated with sanitation conducted

Discussions on specific hygiene needs of women & girls facilitated (see gender pathway)

Hygiene ed in schools facilitated

Participatory assessment of sanitation needs & priorities of women & vulnerable groups

Community management of water & sanitation

Community institutions sustainably manage water & sanitation resources

Preferences & needs of women & marginalised groups are integrated into user plans & implemented

Women & vulnerable groups are fully engaged in water, sanitation & waste mgmt (decision-making, implementation & monitoring of infrastructure investments) (see gender pathway)

Community-based organisations have adequate technical capacity to construct, operate & repair physical infrastructure

Community-based orgs take primarily responsibility for management & governance of all water & sanitation activities

Increased capacity for collective management or governance of water

Comms-based orgs & ex workers trained on interpretation of meteorological forecasts for providing water storage to meet needs during drought periods, & facilitate linkages with appropriate info/tech services for meteo info

Cap building for comms-based orgs & ex agents on technical aspects of physical infrastructure

Sensitize comms on national & local acts, regulations & bylaws on water resources & sanitation

Facilitation of the of implementation of comms awareness campaigns, attitudes & improved practices for safe water, proper sanitation & waste mgmt techniques

Comms-based orgs trained on financial & operational mgmt of water & sanitation systems

Capacity for collection of local meteorological data enhanced

Infrastructure maintenance & mgmt plans (with roles & responsibilities) developed

Water & resource plans which ensure inclusion of marginalised groups developed

Comms participation in linkages with broader infrastructural/mgmt systems where needed

Women's participation in water & waste mgmt promoted

Women-specific meetings organised to hear what their concerns & priorities are (re sanitation)

Discussions facilitated on specific hygiene/ sanitation needs of women (see gender pathway)

Appendix C. List of control localities selected

Table C1. List of 'close by' control localities in Builsa (within 10 km from any MV locality)

	Loc code	AC code	EA	pop10
2	LOC01	AC01	Uwasi	158
8	LOC02	AC02	Banyangsa	788
10	LOC03	AC03	Chansa	150
11	LOC04	AC03	Chansa	1122
43	LOC05	AC01	Naadema	156
46	LOC06	AC04	Jagsa	751
47	LOC07	AC04	Gbedema	631
51	LOC08	AC03	Chansa	200
52	LOC09	AC02	Banyangsa	720
54	LOC10	AC02	Kpikpaluk	270
57	LOC11	AC04	Jiningsa	241
58	LOC12	AC04	Kanjarga	464
59	LOC13	AC04	Jiningsa	764
61	LOC14	AC04	Kanjarga	684
62	LOC15	AC04	Luisa	798
65	LOC16	AC04	Kanjarga	525
66	LOC17	AC04	Kanjarga	125
67	LOC18	AC04	Luisa	611
74	LOC19	AC04	Luisa	633
112	LOC20	AC02	Kpalinsa	713
113	LOC21	AC02	Kpalinsa	695
121	LOC22	AC01	Wiesi	344
132	LOC23	AC03	Zaring- Bulba	578

Table C2. List of 'far away' control localities in Builsa (outside 10 km from any MV locality)

	Loc Code	AC code	EA	pop10
6	LOC24	AC05	Bachonsa	177
14	LOC25	AC06	Chondema	710
15	LOC26	AC06	Achanyeri	184
22	LOC27	AC06	Azug-Yeri	689
32	LOC28	AC05	Doninga	561
60	LOC29	AC07	Daburinsa	356
64	LOC30	AC07	Daburinsa	380
77	LOC31	AC07	Balansa	246
84	LOC32	AC07	Kordema	1786
86	LOC33	AC07	Daborinsa	445
92	LOC34	AC05	Siniensi	611
93	LOC35	AC05	Siniensi	187
94	LOC36	AC07	Kordema	561
95	LOC37	AC05	Siniensi	69
96	LOC38	AC08	Siniensi	427
97	LOC39	AC05	Siniensi	605
98	LOC40	AC05	Siniensi	967
99	LOC41	AC07	Kordema	601
100	LOC42	AC08	Sinyangsa	522
101	LOC43	AC08	Sinyangsa	715
103	LOC44	AC08	Wubilinsa	74
104	LOC45	AC08	Yisobsa	343
105	LOC46	AC08	Bandem	181

Table C3. List of 'close by' control localities in West Mamprusi

	idloc	Loc Code	AC code	AE	Pop10
2	6	LOC47	AC09	Arigu-bisigu	147
8	96	LOC48	AC10	Yizesi	637
14	24	LOC49	AC09		333
21	34	LOC50	AC11	Kurugu	429
23	39	LOC51	AC12		745
35	63	LOC52	AC10	Tantala	2080
41	78	LOC53	AC11	Kurugu	642
44	83	LOC54	AC12	Loagri No. 1	2659
63	1236	LOC55	AC12		1412
84	161	LOC56	AC10	Tantala	794

Table C4. list of 'far away' control localities in West Mamprusi

	Loc Code	AC code	AE	pop10
1	LOC57	AC13	Janga -Fon	216
22	LOC58	AC14		449
24	LOC59	AC15	Nasia	212
34	LOC60	Ac13	kpatorigu	1571
37	LOC61	AC14	Kobore	2687
39	LOC62	AC14	Kubugu	1215
51	LOC63	AC15	Kparigu	688
54	LOC64	AC16	Nayoku	1852
60	LOC65	AC15	Selinvoya	1485
78	LOC66	AC16	Zangum / Zangu-Vuga	2523
81	LOC67	AC14		1682

Table C5: Builsa: Statistical difference between matched and unmatched localities and bias reduction rates (near localities)

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
pop10	Unmatched	492.48	512.03	-8.1		-0.30	0.765
	Matched	492.48	500.83	-3.4	57.3	-0.12	0.908
dpop	Unmatched	.01599	.03152	-29.6		-1.08	0.285
	Matched	.01599	.01396	3.9	86.9	0.16	0.873
dens00	Unmatched	17.473	13.249	90.5		3.47	0.001
	Matched	17.473	14.365	66.6	26.4	2.29	0.027
CHPS	Unmatched	.04348	.14706	-35.2		-1.25	0.218
	Matched	.04348	.04348	0.0	100.0	-0.00	1.000
prim	Unmatched	.43478	.35294	16.5		0.61	0.542
	Matched	.43478	.30435	26.3	-59.4	0.90	0.371
JSS	Unmatched	.04348	.11765	-27.0		-0.96	0.340
	Matched	.04348	.04348	0.0	100.0	-0.00	1.000
JSSdist	Unmatched	6.0783	3.1118	87.9		3.42	0.001
	Matched	6.0783	3.7957	67.7	23.1	2.26	0.029
Mhh	Unmatched	10.795	4.9294	109.6		4.21	0.000
	Matched	10.795	6.2913	84.1	23.2	2.80	0.008
wells	Unmatched	1.3565	1.7647	-32.2		-1.16	0.251
	Matched	1.3565	1.5652	-16.5	48.9	-0.59	0.557

Table C6: Builsa: Statistical difference between matched and unmatched localities and bias reduction rates (far away localities)

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
pop10	Unmatched	492.48	582.64	-23.6		-0.87	0.388
	Matched	492.48	629.35	-35.8	-51.8	-1.05	0.300
dpop	Unmatched	.01599	.03005	-24.4		-0.95	0.344
	Matched	.01599	.02952	-23.4	3.8	-0.69	0.496
dens00	Unmatched	17.473	11.979	123.0		6.00	0.000
	Matched	17.473	14.165	74.1	39.8	2.57	0.014
CHPS	Unmatched	.04348	.1625	-39.5		-1.47	0.145
	Matched	.04348	0	14.4	63.5	1.00	0.323
prim	Unmatched	.43478	.4375	-0.5		-0.02	0.982
	Matched	.43478	.43478	0.0	100.0	0.00	1.000
JSS	Unmatched	.04348	.15	-36.3		-1.35	0.179
	Matched	.04348	.04348	0.0	100.0	0.00	1.000
JSSdist	Unmatched	6.0783	4.3725	41.7		1.75	0.082
	Matched	6.0783	5.4043	16.5	60.5	0.55	0.585
Mhh	Unmatched	10.795	4.9062	116.7		5.87	0.000
	Matched	10.795	6.3	89.1	23.7	2.98	0.005
wells	Unmatched	1.3565	2.25	-66.5		-2.56	0.012
	Matched	1.3565	1.4783	-9.1	86.4	-0.33	0.739

Table C7: West Mamprusi: Statistical difference between matched and unmatched localities and bias reduction rates (near localities)

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
pop10	Unmatched	1237.1	945.82	34.5		1.08	0.286
	Matched	1237.1	1017.9	25.9	24.8	0.61	0.546
dpop	Unmatched	.04392	.03418	32.7		0.89	0.380
	Matched	.04392	.0337	34.3	-4.9	0.88	0.391
builsa	Unmatched	.18182	.11364	18.7		0.60	0.553
	Matched	.18182	.27273	-24.9	-33.3	-0.49	0.631
mamprusi	Unmatched	.72727	.70455	4.9		0.15	0.885
	Matched	.72727	.63636	19.6	-300.0	0.44	0.666
iland	Unmatched	2.5	2.7841	-4.6		-0.11	0.916
	Matched	2.5	.45455	32.8	-620.0	6.71	0.000
wells	Unmatched	11.455	6.7045	53.2		1.51	0.137
	Matched	11.455	11.273	2.0	96.2	0.03	0.975
prim10	Unmatched	1.0909	1.2727	-47.4		-1.26	0.211
	Matched	1.0909	1.1818	-23.7	50.0	-0.60	0.557
JSS10	Unmatched	1.3636	1.75	-81.8		-2.54	0.014
	Matched	1.3636	1.5455	-38.5	52.9	-0.83	0.416
lnjssdist	Unmatched	.87489	1.1209	-18.9		-0.54	0.590
	Matched	.87489	.98741	-8.6	54.3	-0.19	0.850
anyhealth	Unmatched	.09091	.11364	-7.3		-0.21	0.833
	Matched	.09091	.09091	0.0	100.0	-0.00	1.000
market	Unmatched	.63636	.22727	87.8		2.76	0.008
	Matched	.63636	.45455	39.0	55.6	0.83	0.416
floods	Unmatched	3.1818	2.5682	36.3		0.99	0.328
	Matched	3.1818	3.3636	-10.8	70.4	-0.25	0.802
droughts	Unmatched	2.0909	2.3864	-19.7		-0.51	0.613
	Matched	2.0909	2.3636	-18.2	7.7	-0.38	0.705
maize	Unmatched	.81818	.97727	-52.1		-2.13	0.038
	Matched	.81818	.90909	-29.8	42.9	-0.60	0.557
rice	Unmatched	.36364	.45455	-18.0		-0.54	0.595
	Matched	.36364	.45455	-18.0	0.0	-0.42	0.682
millet	Unmatched	.63636	.65909	-4.6		-0.14	0.890
	Matched	.63636	.72727	-18.5	-300.0	-0.44	0.666
fishing	Unmatched	.54545	.38636	31.3		0.95	0.348
	Matched	.54545	.45455	17.9	42.9	0.41	0.687
trading	Unmatched	.90909	.77273	37.1		1.00	0.321
	Matched	.90909	1	-24.7	33.3	-1.00	0.329
crafts	Unmatched	.27273	.29545	-4.9		-0.15	0.885
	Matched	.27273	.45455	-39.2	-700.0	-0.86	0.400

Table C8: West Mamprusi: Statistical difference between matched and unmatched localities and bias reduction rates (far away localities)

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
pop10	Unmatched	1237.1	1111.4	14.2		0.43	0.666
	Matched	1237.1	1252.9	-1.8	87.4	-0.04	0.971
dpop	Unmatched	.04392	.04783	-14.9		-0.43	0.669
	Matched	.04392	.05147	-28.8	-93.4	-0.69	0.499
builsa	Unmatched	.18182	0	63.6		3.28	0.002
	Matched	.18182	0	63.6	0.0	1.49	0.152
mamprusi	Unmatched	.72727	.86	-32.1		-1.07	0.289
	Matched	.72727	.72727	0.0	100.0	0.00	1.000
iland	Unmatched	2.5	10.55	-42.1		-0.98	0.331
	Matched	2.5	16.591	-73.6	-75.0	-1.29	0.213
wells	Unmatched	11.455	8.86	29.3		0.84	0.403
	Matched	11.455	14.182	-30.8	-5.1	-0.51	0.613
prim10	Unmatched	1.0909	1.16	-20.5		-0.58	0.566
	Matched	1.0909	1.0909	0.0	100.0	0.00	1.000
JSS10	Unmatched	1.3636	1.58	-43.1		-1.30	0.199
	Matched	1.3636	1.4545	-18.1	58.0	-0.42	0.682
lnjssdist	Unmatched	.87489	.83611	3.1		0.09	0.925
	Matched	.87489	.88264	-0.6	80.0	-0.01	0.988
anyhealth	Unmatched	.09091	.42	-79.9		-2.10	0.040
	Matched	.09091	.45455	-88.3	-10.5	-2.00	0.059
market	Unmatched	.63636	.12	121.3		4.26	0.000
	Matched	.63636	.45455	42.7	64.8	0.83	0.416
floods	Unmatched	3.1818	1.88	89.0		2.60	0.012
	Matched	3.1818	2.6364	37.3	58.1	0.82	0.420
droughts	Unmatched	2.0909	2.2	-9.6		-0.27	0.786
	Matched	2.0909	2.2727	-15.9	-66.7	-0.35	0.727
maize	Unmatched	.81818	.9	-22.9		-0.76	0.449
	Matched	.81818	.54545	76.3	-233.3	1.37	0.186
rice	Unmatched	.36364	.4	-7.3		-0.22	0.827
	Matched	.36364	.27273	18.2	-150.0	0.44	0.666
millet	Unmatched	.63636	.5	27.0		0.81	0.421
	Matched	.63636	.54545	18.0	33.3	0.42	0.682
fishing	Unmatched	.54545	.46	16.7		0.51	0.614
	Matched	.54545	.45455	17.7	-6.4	0.41	0.687
trading	Unmatched	.90909	.78	35.4		0.97	0.338
	Matched	.90909	.81818	24.9	29.6	0.60	0.557
crafts	Unmatched	.27273	.3	-5.9		-0.18	0.860
	Matched	.27273	.36364	-19.5	-233.3	-0.44	0.666

Table C9: Differences between close-by and faraway control villages in Builsa

Variable	Close-by villages	Faraway villages	p-value
Population 2010	509	517	0.926
Population growth	0.013	0.031	0.303
People per house	14.8	13.5	0.221
CHPS (0/1)	0.04	0.00	0.323
Primary school (0/1)	0.43	0.48	0.773
Distance to JSS	3.5	5.5	0.077*
Distance to market	6.1	7.1	0.441
Number of wells	1.7	1.5	0.544

Appendix D. References

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

Almond, D. and Currie, J. (2011) Human Capital Development before Age Five, in O Ashenfelter and D Card eds. *Handbook of Labor Economics*, Volume 4b, North Holland Elsevier Press.

Anderson, C. L., Dietz, M., Gordon, A. and Klawitte, M. (2004) *Discount Rates in Vietnam*, Economic Development and Cultural Change, 52(4), 873-87.

Ashraf, N., D. Karlan, et al. (2006). "Tying Odisseus to the Mast: Evidence from a Commitment Saving Product in the Philippines." The Quarterly Journal of Economics: 635-672.

Attanasio, O. and B. Augsburg (2011). "Subjective Expectations and Income Processes in Rural India." UCL, mimeo.

Azariadis, C. and J. Stachurski (2005). Poverty Traps. Handbook of Economic Growth. P. Aghion and S. N. Durlauf. Dordrecht, North Holland.

Barrett, C. B., P. P. Marenya, et al. (2006). "Welfare dynamics in rural Kenya and Madagascar." Journal of Development Studies 42(2): 248-277.

Bauer, M. and Chytilova, J. (2010) *The Impact of Education on Subjective Discount Rate in Ugandan Villages*, Economic Development and Cultural Change, 58(4), 643-69.

Bertrand, Duflo and Mullainathan, 2005, How Much Should we Trust Difference in Difference Estimates?

Bourguignon, F., F. H. G. Ferreira, et al. (2007). "Inequality of opportunity in Brazil." Review of Income and Wealth 53(4): 585-618.

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

Caselli, F. and W. J. Coleman (2001). "The US structural transformation and regional convergence: A reinterpretation." Journal of Political Economy 109(3): 584-616.

Chen, S. H., R. Mu, et al. (2009). "Are there lasting impacts of aid to poor areas?" Journal of Public Economics 93(3-4): 512-528.

Clemens, M. and Demombynes, G. (2010), When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages, World Bank Policy Research Working Paper 5477, USA.

Coulombe, H. and Quentin Wodon (2007) *Poverty, Livelihood, and Acess to Basic Services in Ghana*, The World Bank, Ghana Cem: Meeting the Challenge of Accelerated and Shared Growth, Partial and Prelimiary Draft for Review.

Dhaliwal I, Duflo E, Glennerster R, Tulloch C, (2011). Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education, JPAL, Working Paper.

Dixit, A. (1973). Models of Dual Economies. Models of Economic Growth: proceedings of a conference held by the International Economic Association at Jerusalem. J. A. Mirrlees and N. H. Stern. London, Macmillan.

Frederick, S., G. Loewenstein, et al. (2002). "Time Discounting and Time Preference: A Critical Review."

Journal of Economic Literature XL: 351-401.

Jalan, J. and M. Ravallion (1998). "Are there dynamic gains from a poor-area development program?" Journal of Public Economics 67(1): 65-85.

Jimenez E and Patrinos HA, (2008). Can Cost-Benefit Analysis Guide Education Policy in Developing Countries? World Bank, Policy Research Working Paper No. 4568;

Lybbert, T. J., C. B. Barrett, et al. (2004). "Stochastic wealth dynamics and risk management among a poor population." Economic Journal 114(498): 750-777.

Manski, C. (2009). "Vaccination with Partial Knowledge of External Effeectiveness." PNAS 107(9): 3953-3960.

Mazumder, B. (2005). The Apple Falls Even closer to the Tree than We Thought: New and Revised Estimates of the Intergenerational Inheritance of Earnings. Unequal Chances: Family Background and Economic Success. S. Bowles, H. Gintis and M. Osborne Groves. Princeton, NJ, Princeton University Press.

McKenzie, D. J. and C. Woodruff (2006). "Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises." Economic Development and Cultural Change 55(1): 3-42.

Millennium Project (2005) I*nvesting in Development. A Practical Plan to Achieve the Millennium Development Goals*, Report to the UN Secretary General.

Millennium Project (2006) *Q/A on the Millennium Villages*, November 2006, UN Secreatary-General and the UN Development Group.

Millennium Promise (2010), Harvests of Development in Rural Africa: The Millennium Villages after Three Years.

Mensah-Homiah, J., Afram, S., Ohmeng-Dapaah, S., Kim Huynh, U., Radunsky, A. and Rose, A. (2011), *Bonsaaso Ghana: Three Year Review*, March 2011, Millennium Villages Promise.

Pritchett, L. (1997). "Divergence, big time." Journal of Economic Perspectives 11(3): 3-17.

Pronyk, P. et al (2012) The effect of an integrated multi-sector model for achieving the Millenium Development Goals and improving child survival in rural sub-Saharan Africa: a non-randomised controlled assessment, in *The Lancet* 379 (9827).

Ravallion, M. and S. H. Chen (2005). "Hidden impact? Household saving in response to a poor-area development project." Journal of Public Economics 89(11-12): 2183-2204.

Ray, D. (1998). Development Economics. Princeton, Princeton University Press.

Sachs, J.D.; J.W. Mcarthur; G. Shmidt-Traub; M. Kruk; C. Bahadur; M. Faye and G. McCord. 2004. "Ending Africa's Poverty Trap." Brookings Papers on Economic Activity, 1(2004), 117-240.

Sadoulet E. And A. De Janvry (1995), *Quantitative Development Analysis*, Baltimore: John Hopkins University Press

Tan-Torres Edejer T, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans DB and Murray CJL. (2003). Who Cost-effectiveness Analysis, Geneva: WHO.

Tarozzi, A. and Deaton, A. (2009) *Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas,* The Review of Economics and Statistics, 91(4), 773-92.

Taylor, E. (2012) A methodology for local economy-wide impact evaluation of cash transfers, in by J. Edward Taylor and Mateusz Filipski eds. *Beyond Experiments: Simulation Methods for Impact Evaluation*.

Voors, M. J., Nillesen, E., Verwimp, P., Bulte, E., Lensink, R. and Van Soest, D.P. (2012) *Violent Conflict and Behavior: A Field Experiment in Burundi*, American Economic Review, 102(2), 941-64.

Weiss, C. (1998) *Evaluation: methods for studying programs and policies*, Prentice Hall, New York, USA.

White, H. (2009) *Theory-based impact evaluation: Principles and Perspectives*, June 2009, Working Paper 3, International Initiative for Impact Evaluation, New Dehli, India.

Yotopoulos P.A. and L.J. Lau (1979) Introduction, *Food Research Studies*, Vol. XVII, No. 1

Appendix E. Evaluation Survey and Other Instruments

PART 1: QUANTITATIVE SURVEY INSTRUMENTS

PART 2: QUALITATIVE METHODS AND TOOLS

PART 3: QUALITY ASSURANCE PROTOCOLS AND TEMPLATES