Impact Evaluation of Water Supply Activity: Baseline Report

PRESENTED TO:

Government of Ghana
Millennium Development Authority
Office of Monitoring and Evaluation
4th Floor Heritage Tower, 6th Avenue
Ridge West, Accra, Ghana

PRESENTED BY:

Raymond Struyk, George Caldwell, John Felkner, Samuel Haddaway, David Hemson, and Kareem Kysia
NORC at the University of Chicago
4350 East-West Highway
Bethesda, MD 20814
Tel. (001)(301)634-9300

MiDA Contract: 4101106-01
NORC Project Number: 6638

APRIL 15, 2011

NORC at the UNIVERSITY of CHICAGO

This page left intentionally blank

Executive Summary

The Millennium Development Authority of Ghana (MiDA) has allocated some of the funds received from its Compact with the U.S. Millennium Challenge Corporation (MCC) to improve critical infrastructure in certain communities located in agricultural areas in which MiDA operates. The Rural Development/Community Services Project is designed to enhance the impact and sustainability of MiDA's Agriculture Project by providing necessary social infrastructure and service delivery. Communities in the Districts targeted for the program have poor access to basic community services such as potable water, sanitation, and schools and also lack electric power to drive rural industry or domestic use.

Water supply is one of the infrastructure projects for which MiDA is financing investments under the Community Services Water Supply Activity (WSA). It is being implemented in the Afram Basin Zone, Southern Horticultural Belt, and the Northern Agricultural Zone. A total of 137 communities were selected for water system upgrades and construction is underway at this time.

MiDA has contracted NORC to design a rigorous evaluation of the impact of the improved water supply services and to conduct the baseline survey. Of particular interest are effects on household health outcomes, time savings, and income levels.

This report presents the design of the impact evaluation, the sampling plan, plans followed in conducting the baseline survey and the questionnaires administered in treatment and control areas to households and the community survey administered to a knowledge local person about conditions generally in the community. The surveys completed interviews with 1,200 households in 100 communities, 50 treatment and 50 control communities and with a knowledgeable civic leader in each community. A documented data set has been prepared.

There many patterns in the baseline data of policy interest, the most important of which we now highlight. The household survey made careful inquiries about sickness in a household during the two weeks before the interview took place—a recall period widely recognized as quite accurate. The overall (total) incidence of illness or injury reported in the GLSS5+ and the water surveys is quite similar, at 13 percent for GLSS5+ and about 15 percent for the household water survey. Of particular interest is the incidence of illness for children under 5 years old because of the high mortality rates for these children associated with water borne disease. The incidence is high at 18 percent of these young children in the GLSS5+ to 26 percent among infants and toddlers in treatment households.

The incidence of persons suffering from diarrhea during the same time period is about 4 percent for children four and under, which is about three times higher than the overall incidence among persons in the surveyed households.

Poor hygiene is a common cause of stomach illnesses, and the survey asked a series of questions on this topic. Overall practices are not very strong. Households have a low incidence of facilities to improve hygiene such as a place to wash one's hands or a soak-away pit. In only about one-third of households did anyone attend a hygiene promotion event in the past year. When asked about how the youngest child's stools are disposed, only about 40 percent reported putting them in a toilet facility, 29 percent went into the trash. The respondent's own hygienic practices are mixed: While nearly two-thirds said they always wash their hands with soap after using the toilet, only about 40 percent reported always washing their hands before eating or preparing food.

Turning to the type and quality of dwellings occupied by the households in the surveyed communities, one finds that a sizable majority of households in the surveyed communities —about 75 percent--live in rooms within a larger structure that is shared with other room owning or renting occupants. Most of these dwellings are compound houses. The next most frequent arrangement is the occupancy of multiple huts within a compound. The patterns are broadly the same in data from the GLSS5+ and the water survey, with differences likely attributable to the GLSS5+ data being for all areas, not just rural zones.

Eighty percent of households surveyed own the unit in which they live. About 10 percent are rent paying occupants and 5 percent occupy units rent free.

There are sharp differences in the quality of dwellings as indicated by the durability of the materials from which they are constructed. Households are evenly balanced between those having units with walls made of mud or mud bricks and those with walls constructed from cement or sandcrete blocks. In contrast, the incidence of durable materials is greater for floors and roofs. Floors are of particular interest because analyses have shown that cement floors have very large positive effects on toddler health and children's school performance. In the communities surveyed 78 percent of the dwellings occupied have cement or concrete floors. Because of the Ghana's heavy rains, a strong roof is also a important contributor to good health. Here the incidence of strong materials, corrugated iron sheets, at 60 percent of units is less than for durable flooring. Clearly the analysis of health improvements associated with improved water services will have to control for the quality of respondents' walls, floors, and roofs.

Sixty percent of respondents in both treatment and control communities get their water from wells broadly defined. But there is considerable diversity between the treatment and control communities for other sources, with households in control communities more often using pipe-borne water (15 to 6 percent for treatment communities) and less often using rain and surface water which is often of low quality (21 to 33 percent for treatment communities). These patterns confirm that MiDA is targeting communities with comparatively dire water problems with its investments. Regarding the toilet

facilities used by households in the water survey, only 38 and 33 percent, respectively, of those in control and treatment communities, used improved facilities. About 30 percent in both community types use no type of facility whatsoever.

The analysis focuses on water sources used in the dry season because water investments will yield their greatest benefits during this season. Several sources are used significantly in the dry season only: About 29 percent of households use water from someone else's yard only during the dry season and 48 percent only use water from tanker trucks then.

Respondents were also asked if the water from each source is safe to drink in the dry season. Three sources are rated by 38-45 percent of households as unsafe: unprotected dug wells, tanker trucks, and surface water, e.g., a lake. On the other hand, tube wells and public standpipes were rated safe by nearly everyone.

The survey obtained detailed data on the time in minutes devoted to collecting water. Information was obtained on time spent by each water fetcher in each of the three distinct steps in the acquisition process: going to/from the water sources, time waiting in line to collect the water, and the time to collect the water, e.g., to draw water from a well.

Water sources fall into three groups based on the total time required.

- There are six sources with high time requirements per trip, i.e., times approaching one hour. These are the first seven sources in the table with the exception of a public standpipe which has lower times. Average time to go to/from a source is 16-30 minutes, while waiting time ranges from 21 to 37 minutes. Actually obtaining the water averages 4 to 6.5 minutes.
- The second group has three sources with total average times in the 5 to 18 minute range: collected rainwater, bottled water and tanker trucks. Times going to/from sources are very short except for tanker trucks (11 minutes).
- The final group of three sources is in an intermediate position for total average acquisition time: public standpipes, carts with a small tank or drum, and surface water. Total average time ranges for 25 to 39 minutes.

The typical person collecting water makes 12-15 trips per week for sources where significant time is required. But the trip numbers for the three low total time sources and the cart-with-drum source are in the 6-10 range, perhaps indicating these are supplemental rather than primary sources.

When one aggregates the trips taken by all household members, one gets the total minutes per household per day that households spend getting water from a source. The highest value is 270 minutes per day for water from an unprotected spring. The mean values for four sources are around

200 minutes per day: getting water from someone else's plot, a tube well/borehole, an unprotected well, and a protected spring. The values for collected rainwater and bottled water are dramatically lower. It is important to look at the distribution of household over the various time categories because they show that while the mean values are often high, large minorities and often the majority of households using a source spent less than 60 minutes a day at the task.

Because essentially all households use multiple water sources, one anticipates that total times for all sources used will generally be larger than those for individual sources. And this is the case. All household members together are spending around four hours a day (240 minutes) collecting water. There is very little difference between the time for collecting water for drinking and cooking purposes and all purposes, indicating that most effort is going to meet essential water requirements. But the distribution of time spent among households is also very important and informative. About 20 percent spend less than 30 minutes a day fetching water, and about half spend under two hours a day. Rather shockingly, about 25 percent have household members devoting over six hours a day to acquiring water.

So what do households pay in a month for water from all the sources they access? Total spending is computed by weighting the price of water from each source used a household by the share of total water consumption accounted for by the same source. Households in control communities pay GHc 27 and those in treatment communities GHc 48 on average. Because fewer households in treatment communities pay for water from any source, the result indicates that those who do pay are paying higher prices than those in control communities. From other data we infer that households in treatment communities may have been more often purchasing water from unimproved sources.

The discussion of water collection and cost has referred to multiple water sources used by a household. The typical household accesses 2-3 water sources on a regular basis. About 15 percent use 4 or more sources. Once again the patterns for treatment and control community types are essentially identical.

Acknowledgements

The work reported here has benefitted substantially from our interaction with MiDA's Office of Monitoring and Evaluation. In particular, Mrs. Abigail Abandoh-Sam and Albert K. Nyarko provided consistent advice and guidance. Mr. Nyarko was instrumental in the team acquiring data needed for matching treatment and control localities. MiDA's Office of Monitoring and Evaluation and Jigar Bhatt from the Millennium Challenge Corporation provided extremely valuable comments on the draft of this report.

Table of Contents

1. Introduction 1
1.1 The Water Supply Activity 1
1.2 Hypotheses 4
1.3 Report Organization 5
2. Evaluation Design 7
2.1 General Considerations 7
2.2 Description of the Process for Selecting Treatment Communities for the Program and the Corresponding Population of Potential Control Communities 12
3. Questionnaire Development 25
4. Data Collection 29
4.1 Development of Operational Plan and Manuals 29
4.2 Institutional Review Board (IRB) 30
4.3 Recruitment and Training of Field Staff 30
4.4 Implementing Data Collection 34
4.5 Supervision and Quality Control 35
4.6 Data Processing 36
4.7 Results 36
5. Baseline Data and Analysis 39
Introduction 39
Information from the Household Survey 41
Information from the Community Survey 77
References 83
Annex A: Questionnaires 85
Annex B: Tables for Selection of Control Communities 113
Annex C: Procedures for Identifying and Replacing Outliers in Water Data 139

Table of Tables

Table 1.1: Alternative Water Supply Investments Funded by MiDA 3

Table 1.2: Effects of Improved Water Supply in Rural Communities 4

Table 4.1: Training Schedule 33

Table 4.2: Household Survey Results 37

Table 5.1: Definitions of Improved and Unimproved Water and Sanitation Facilities 40

Table 5.2: Average age of household head, by sex, MiDA Zone and locality (%) 41

Table 5.3: Distribution of households, by adult composition, locality and presence of children (%) 42

Table 5.4: Population aged 15 years and older, by educational attainment and sex (%) 42

Table 5.5: Incidence of people suffering from illness or injury during the previous two weeks, by age group and sex (%) 43

Table 5.6: Incidence of people suffering from diarrhea during the previous two weeks, by age group and sex (%) 43

Table 5.7: Proportion of people suffering from an illness or injury who had to stop their usual activity two weeks preceding the interview, by locality, age group, and sex (%) 44

Table 5.8: Care received by sick persons 45

Table 5.9: Household hygiene 46

Table 5.10: Respondent Hygiene 47

Table 5.11: Current economic activity rate, by sex, age group and locality (%) 48

Table 5.12: Number of workers per household and types of occupation (%) 48

Table 5.13: Main Occupation of Working Household Members (percent of those working) 49

Table 5.14: Type of employer for the currently employed population aged 15-64 years, by locality and sex (%) 50

Table 5.15: Payments for work in main and secondary occupations 51

Table 5.16: Monthly incomes per working household member and per household 52

Table 5.17: Durable goods ownership by households 52

Table 5.18: Households, by type of dwelling (%) 54

Table 5.19: Households, by present occupancy status (%) 54

Table 5.20: Households by number of rooms occupied (%) 54

Table 5.21: Main construction materials used by households (%) 55

Table 5.22: Households by main source of water for drinking and cooking (%) 56

Table 5.23: Households by main source of water for general uses (%) 57

Table 5.24: Households by type of toilet used (%) 58

Table 5.25: Means of building and paying for toilets, Households in MiDA survey 59

Table 5.26: Water use and treatment, for water sources used for cooking and drinking during the dry season or both seasons 66

Table 5.27: Water use and treatment, for water sources used for general purposes during the dry season or both seasons 68

Table 5.28: Time to collect water during the dry season 70

Table 5.29.1: Total household time spent fetching water per day (minutes) 71

Table 5.29.2: Total household time spent fetching water per day per capita 71

(minutes (1)) 71

Table 5.30: Water volume and monthly cost, all water sources 72

Table 5.31: Summary of Household Water Consumption and Incidence of Water Payments 73

Table 5.32: Water volume and monthly cost, water sources used mainly for cooking/drinking 74

Table 5.33: Incidence and Amount of Water Payments by Households in Control and Treatment Communities 74

Table 5.34: Total household water cost, All Households, per month (GHc) 75

Table 5.35: Total household water cost, All Households, per capita per month (GHc) 75

Table 5.36: Total household water volume collected per day 76

Table 5.37: Total household water collected per day per capita 76

Table 5.38: Number of sources used per household 76

Table 5.39: Number of Households Using Various Combinations of Water Sources 76

Table 5.40: Indicators of Community Conditions 78

Table 5.41: Types of water sources used in communities 79

Table 5.42: Characteristics of communities' most important water source 80

Table 5.43: Interviewer's direct observations about the community 81

Table 5.44: Schools in communities, by school type, sponsor, and number 82

1. Introduction

The Millennium Development Authority of Ghana (MiDA) has allocated some of the funds received from its Compact with the U.S. Millennium Challenge Corporation (MCC) to improve critical infrastructure in certain communities located in agricultural areas in which MiDA operates. The Rural Development/Community Services Project is designed to enhance the impact and sustainability of MiDA's Agriculture Project by providing necessary social infrastructure and service delivery. Communities in the Districts targeted for the program have poor access to basic community services such as potable water, sanitation, and schools and also lack electric power to drive rural industry or domestic use.

Water supply is one of the infrastructure projects for which MiDA is financing investments under Community Services Water Supply Activity (WSA). It is being implemented in the Afram Basin Zone, Southern Horticultural Belt, and the Northern Agricultural Zone. A total of 137 communities have been selected for water system upgrades.

MiDA has contracted NORC to design a rigorous evaluation of the impact of the improved water supply services and to conduct the baseline survey. Of particular interest are effects on household health outcomes, time savings, and income levels. This report presents the design of the impact evaluation, the sampling plan, plans followed in conducting the baseline survey and the questionnaires administered in treatment and control areas to households and to a knowledge local person about community conditions generally.

The balance of this chapter outlines the water supply program and states the hypotheses to be tested in the evaluation.

1.1 The Water Supply Activity

It is useful to distinguish among four water supply solutions within the overall water supply activity. All solutions include a hygiene education component. The water supply at the delivery source in all cases meets official Ghana water standards set by the Community Water and Sanitation Agency (CWSA). The quantity of water produced is sufficient to meet basic domestic needs for cooking and consumption.[1] In the following, the four solutions are outlined and then the local administration of

[1] According to van Koppen et al. (2009, S78), basic domestic requirements are satisfied with 5-20 liters per capita per day. According to MiDA community services consultants, the borehole wells with a hand pump being installed by the project can be used continuously for six hours before needing a three hour recovery period and can produce at least 13 liters per minute. Assuming that the pump is used the equivalent of six full-time hours each day and the borehole serves 300 persons, then per capita daily production is at least 15.6 liters. While this is well within the range to supply basic domestic requirements it

the water supply is described. The selection of beneficiary communities is discussed in Chapter 2 on the evaluation design and sampling.

Solutions. The four solutions are summarized in Table 1.1. *Boreholes* are drilled and fitted with hand pumps for small communities of a population of less than 2,000. The small town extension solution is employed based on the criterion that population exceeds 2,000 persons. If the inadequacy of water supplies in village through boreholes was identified after the allocation of funds among beneficiary communities, a small town extension is substituted but within the budget for boreholes, resulting in fewer supply points than would have been available under the borehole solution.

The *small town water system* comprise a high yielding borehole mechanized to lift water into an overhead tank and the water is then distributed under gravity to standpipes strategically located in the community such that the maximum distance a person has to walk to fetch water should be 500 meters. The small town water system can also be dependent on surface water where high yielding boreholes are not feasible but existing streams and water bodies are available throughout the year. *Small town piped extensions* are extensions of existing pipe networks from nearby communities with enough capacity to add-on nearby communities.

The Tamale water extension in the Northern Zone is an example of a piped extension. However, it is a piped extension to communities close to Tamale and employs an Urban System and not a small town system. The water source is the rehabilitated Tamale water works. In this case water is treated before it is transmitted because the area has traditionally suffered from a guinea worm epidemic. The facility serves the Tamale metropolitan area and two near-by districts—Tolon-Kumbungu and Savelugu-Nanton. TheTamale metro area is not included in the water supply component. Because of budget constraints, the Tamale extensions have been sharply reduced from the original plan from 13 to 4 communities and from 11 to 9 in the near-by communities in Savelugu-Nanton.

falls short of the volume needed for multiple-use water services as indicated by van Koppen *et al., ibid.* WHO estimates minimum water requirements for drinking and cooking at 20 liters per capita per day.

Table 1.1: Alternative Water Supply Investments Funded by MiDA

Name	Community Size (population)	Comments
Boreholes	Under 2,000	Standard: 1 borehole per 300 persons
Small town	2,000-5,000	1standpipe with two spouts, 1 spout per 300 persons; water pumped to holding tank and distributed through gravity pipe system
Pipe extension	2,000-5,000	Water is piped from an existing water treatment system of Ghana Water Company Limited (GWCL) in a community or town (within 1-6 km) to the target community and there distributed as in the small town solution.
Tamale water extension	2,000-5,000	Water piped from the large water treatment facility in Tamale to near-by communities; distribution is similar to the small town solutions

Administration. Each community forms a special administrative body responsible for the water-related issues. In smaller communities it is called the Water and Sanitation Committee and in towns the Water and Sanitation Development Board. Hereafter we refer to this body as "the Board." The Board is responsible for maintaining the water supply system, setting tariffs and collecting them, and the education of population in water-related hygiene.

With respect to hygiene, Board members are being trained by MiDA consultants in a hygiene course. The Board is responsible for delivering this information to all households in the community and to remind citizens periodically of good hygienic practices.

The GWCL and the CWSA have established guidelines on setting tariffs and collection methods, but each Board holds discussions with the community and agrees on local rules. In the end, however, the community benefitting from the GWC extension must pay the bills presented by GWC or risk water being shut off. In practice most communities authorize a vendor to collect a fee at each distribution point for each vessel filled, i.e., a "pay as you fetch" system. In other cases the Board may decide to make a collection when a GWC bill is received. On this basis, it seems likely the price for water reported by households in treatment communities will vary significantly.

1.2 Hypotheses

MiDA is interested in documenting project effects in several areas of greater availability of water meeting minimum standards through the impact evaluation. Many evaluations are more narrowly focused, with primary attention often going to health effects, especially the reduction in the incidence of diarrhea in young children (Waddington and Snilstveit, 2009). This priority is understandable, since diarrhea is responsible for an estimated 21 percent of fatalities of under-fives in developing countries or 2.5 million deaths per year (Kosek at al., 2003).

In Table 1.2 we list the five types of effects to be measured in the evaluation and state the working hypothesis to be tested for each in the evaluation. In every case the hypothesis is that the improvement significantly exceeds that of a similar control group.

Table 1.2: Effects of Improved Water Supply in Rural Communities

Area	Hypothesis
Health	**The incidence of diarrhea will decline, particularly in children 5 years old and younger.**
Time savings	Time devoted to acquiring water will fall significantly. Time freed through more efficient water collections will be shifted significantly to income producing activities
Water prices	The price paid for drinking water will decline significantly, where water has been previously purchased.
Quantity of water consumed	Households will consume a greater quantity of water for domestic purposes
Household welfare	Household consumption expenditures will increase

Improvements in water supply accessibility and quality have been documented to have significant impacts in each of the areas listed.[2] That said, particularly for improvements in health, the gains are far from assured. It is well documented that water that is clean at its source is subject to many opportunities for degradation through the way it is handled and stored before home consumption. Because the MiDA water improvements do not include investments in better sanitation facilities, the opportunity for degradation are likely to be higher than in more comprehensive water, sanitation and hygiene (WSH) programs.

[2]Waddington and Snilstviet (2009) cover health effects in detail; on time savings, see, for example, Aiga and Umenai (2002) and Hutton et al (2006). Water quantity effects are documented in Aiga and Umenai (2002), for example. Price effects for urban areas are reported in UN-Habitat (2003, 114); we have not found estimates for rural programs. Income effects, arising from avoiding costs associated with water-related illness and increased income from the greater work effort thereby permitted can in principle be inferred from willingness-to-pay and hedonic model estimates that are summarized in Pattanayak et al (2010); they also present a comprehensive estimate of the value of an Indian water and sanitation program.

Increases in the quantity of water more readily available to households can have health effects through increased direct consumption, preventing dehydration, and indirectly through water not being recycled within the home for various purposes. It can also have economic effects as when greater amounts of water can be applied to growing vegetables, crops and trees and watering livestock (van Koppen, Smits, and Mikhail, 2009).

Household consumption may increase through three channels. First, the time freed through shorter distances to water sources and shorter queues at sources may be devoted to economic activity ranging from growing vegetables to doing piece-work at home. Second, a lower diarrhea incidence for young children will release time of the caregiver for economic activity. Third, a lower incidence of adult illness from water-related diseases translates into fewer work days lost and increased income.

The evaluation will investigate all of these effects. However, it is worth noting at the outset that the household interviews was limited to 90 minutes on average which restricts the extent to which household welfare effects in particular can be investigated. The specific information to be gathered in the surveys is discussed in Chapter 4.

1.3 Report Organization

The balance of this report consists of six chapters and two annexes. The next chapter presents the evaluation design and Chapter 3 outlines the sampling plan and procedure to match treatment and control communities. Chapter 4 describes the basic principles underlying the development of the questionnaire, and Annex 1 contains the questionnaires. Chapter 5 outlines how the field work was conducted and Chapter 6 presents tabulations of the baseline data.

2. Evaluation Design

2.1 General Considerations

The Community Services' Water Supply Activity provides improved water systems to selected communities (usually villages, but larger administrative areas in some cases) in MiDA program areas. The selected communities are referred to as "treatment units." The basic approach to the impact evaluation is to compare changes over time in variables of interest for the treatment units to changes for similar communities over the same time interval. The non-treatment units to which the treatment units are compared are referred to as "comparison units" or "control units." Although the program is administered to communities, the principal unit of observation and analysis is the household.

Evaluation Design: Pretest-Posttest-with-Matched-Comparison-Group

The evaluation design for this project is a "pretest-posttest-with-matched-comparison-group quasi-experimental design," which consists of four groups: a "treatment" group at two points in time ("before and after," "pretest/posttest," "time 1 and time 2") and a "comparison" (or "control") group at the same two points in time. The "before" time is a time shortly before completion of the water system improvements, and the "after" time is some time after completion of those improvements. The measure of program impact is a "double-difference estimate," which is the difference, between the treatment and comparison groups, of the difference in various outcome measures of interest (e.g., time savings in acquiring water, water consumption and income) before and after the program intervention. Time 1 was the August-September 2011 baseline data collection. Time 2 will occur some time later (the present contract is for the evaluation design and the baseline survey – the timing of the end-line survey is not yet determined). Research suggests that households respond quickly in using improved water supplies. Because the focus in the baseline survey is on water sources used in the dry season, then a dry season should occur between the time when construction is complete and the Time 2 survey.

Population of Control Units

The population of control communities should be as similar as possible overall to the population of treatment communities, with respect to variables that may affect outcomes of interest to the program. For this application, the control population was defined to be communities located within MiDA program areas, considered eligible to receive water improvement services but not receiving such improvements from MiDA or other sources.

Survey Design: Stratified Two-Stage Sampling, with Variable Probabilities of Selection

For the survey, a two-stage survey design was proposed, in which a first-stage sample of communities is selected, and a second-stage sample of households is selected from within each sample community. (The first-stage sample units, or primary sample units (PSUs), are communities. A sample of 12 households from each sample community was selected with the objective of having a post test sample of 10 households for each community after attrition between surveys. This type of design is efficient for surveys covering large geographic areas (i.e., it achieves a high level of precision for estimates of interest, or a high level of power for conducting tests of hypotheses, for sampling effort expended). The sample sizes (numbers of communities and numbers of households within each sample community) were determined to provide a high probability ("power"; 90 percent) for detecting an income increase of 10 percent (between the before and after surveys), and for detecting a 20-percent change in a proportion. The final sample size for the household survey is 1,000. At baseline 1,200 interviews were completed to allow for attrition between survey rounds.

To further improve precision and power, two additional design features were used in addition to multistage sampling: stratification and selection of primary sample units (PSUs; communities) with variable probabilities. Stratification was employed to control the distribution of variables, known prior to the survey, that have an effect on outcomes of interest. Since several stratification variables were involved, the stratification technique used is called "marginal stratification" (as contrasted with "cross-stratification"). Selection of primary sample units with variable probabilities is used for two purposes: to implement the marginal stratification, and to control the probability of selection of the ultimate sample units (households, which are the primary unit of analysis).

Analytical Survey Design

In most program evaluations, it is desired not just to obtain an overall estimate of program impact, but also to estimate the relationship of impact to other variables that affect it. (This is done not only because such relationships are of interest in their own right, but also because when quasi-experimental designs are used it is desirable to adjust impact measures to account for differences in the distributions of the treatment and control groups with respect to variables other than treatment (i.e., to estimate "counterfactuals" or "potential outcomes" by adjusting for "covariates").) Measures of program outcome and impact are "response variables" or "dependent variables," and variables that affect them are "explanatory" or "independent" variables. The type of sample survey used to support impact evaluation assessment is called an "analytical" survey. The goal of this type of survey is to test hypotheses about program impact (e.g., detect a program-caused change of a specified size), and to estimate the relationship of impact to explanatory variables. This type of survey differs very much in nature from the more common "descriptive" survey, which is intended to produce estimates of overall

population characteristics, such as population means, proportions and totals. In order to be able to estimate relationships, it is desirable that the sample be designed so that there is substantial variation ("spread") in important explanatory variables, and low correlation among them. In the present application, we shall achieve these objectives by stratifying the sample on variables, known prior to the survey, that are believed to have an effect on outcome.

Marginal Stratification; Selection with Variable Probabilities

The approach to stratification used in this application (analytical survey design) is different from that used in descriptive sample surveys, because the number of variables of stratification is potentially large. In this application, communities were selected with variable probabilities, where the probabilities are determined so that the expected number of units in each stratum category (or "cell") is as close as possible to a desired allocation (having a reasonable level of spread in important explanatory variables, and a low level of correlation among them), and such that the probability of selection is positive for every unit of the treatment population. With this approach, constraints are imposed on the desired expected number of units in each stratum cell, but because of sampling fluctuations and the nature of the population, the stratum sizes may not be exactly as desired, for any particular selected sample.

The Use of Matching to Reduce Selection Bias and Increase Precision and Power

In the simplest example of a pretest-posttest-with-comparison-group" experimental or quasi-experimental design, all four groups are selected independently. In order to improve the precision of difference estimates and the power of tests of hypotheses about them, it is desirable to match individual units of the "before" and "after" groups, and the "treatment" and "control" groups. This type of design is called a "matched pairs" design. For groups sampled at different times (longitudinally; "before" and "after"), the standard approach to matching is to use "panel" sampling, in which the same households are interviewed at Times 1 and 2, for both the treatment and control groups. For groups sampled at the same time (cross-sectionally), the standard approach is to match individual primary sampling units (communities) in the treatment and control groups. The matching is done taking into account variables (known prior to the survey) that may affect program outcome (i.e., outcomes of interest to the program, such as income, water consumption, or time savings in acquiring water).

Matching to Reduce Selection Bias

The "ideal" pretest-posttest-with-comparison-group design for an impact evaluation study is an experimental design in which the experimental units are selected randomly from an eligible population, and selection for treatment is also done using randomization. Randomized assignment of

treatment assures that the probability distribution of all variables is the same for the treatment and control groups. When randomization is not used for treatment selection, a "selection bias" may be introduced into estimates, and tests of hypothesis may be corrupted. In the present application, randomization was not used to determine inclusion in the program. Because of this, the control group is determined by matching, not randomization. A primary goal of the matching process is to reduce selection bias by making the probability distributions as similar as possible for the treatment and control groups, for all variables (known prior to conducting the survey) that are believed to affect outcome. This type of matching is called "matching on observables," or "ex-ante" matching. Data on the match variables were obtained from MiDA and from vendors of geographic information system (GIS) data.

Matching to Increase Precision and Power

While matching is used to reduce selection bias, it is not the only reason for using it. Matching was used in the present sample design for two reasons – to reduce selection bias that may be introduced because randomization was not used to assign treatment to communities, and to increase precision and power. Matching reduced selection bias by causing the distributions of design variables related to outcome to be similar for the treatment and control samples, as would be the case if random assignment of treatment had been used. It has increased precision through the formation of "matched pairs" of individual treatment and control units. The use of matched pairs increased the precision of difference estimates, such as the double-difference estimate of program impact, over that available from designs based on matched samples, but not individually-matched pairs.

Sample Size Determination; Statistical Power Analysis

Since there are two levels of sampling involved in the present survey, there are two types of sample size: the sample size of the communities (first-stage sample units; primary sample units) and the sample size of households (second-stage sample units; ultimate sample units; elements) within communities. Based on a statistical power analysis, and taking into account budgetary and time constraints, it was decided that a sample of 50 communities would be selected for each group (i.e., 50 treatment communities at Time 1, 50 control communities at Time 1, 50 treatment communities at Time 2 (the same treatment communities as at Time 1), and 50 control communities at Time 2 (the same control communities as at Time 1)). Taking into account the relative costs of sampling communities vs. households, and the likely homogeneity of communities relative to outcome variables of interest (as measured by the intra-community correlation coefficient), it was decided to select samples of 10 households from within each sample community.

Handling of Nonresponse

With the adopted sampling approach, a sample of treatment units was selected (according to the stratum allocations), and two control units were matched to each selected treatment sample unit. To allow for nonresponse, two features were incorporated into the design: (1) the sample size was increased from the desired number of 50 communities for each group (treatment before, control before) to 60 for each group; (2) for each treatment unit in the sample two matching control units were selected. Hence, the complete sample consists of a set of 60 matched triplets, of which the first 50 acceptable comprise the desired sample. If for any reason it is not possible to observe a particular treatment community in a matched triplet, that triplet is discarded and a replacement matched triplet is selected. For controls, the procedure was a little different. If the first-selected control unit was acceptable, then it was used as the matching control. If it was not acceptable, then the second-selected control unit was used. If both controls were unacceptable, then an unused control from the same district is used.

It is noted that while the treatment sample is a probability sample, in this quasi-experimental design the control sample is not. The goal was to match the best possible control unit for each treatment unit. This is done by matching, and does not involve probability sampling. The fact that there are multiple matching control units for each treatment sample unit does not affect the composition of the treatment sample, or affect the probabilities of selection. It does not bias the sample. Replacement of treatment units could bias the sample. For this reason, replacement of treatment units would be done only under extreme circumstances, such as the unit's being out of scope, civil disturbance, medical quarantine, or other extreme emergency.

Selection of Households within Communities

The preceding paragraphs have described the procedure for selecting first-stage sample units (communities). The selection of a random sample of households from within each sample community was made during the course of the survey, when each sample community was visited. The procedure for selecting a systematic random sample was as follows.

In each community, a **starting point** was selected by the supervisor or field manager. This could be a structure like a church, mosque, school, hospital, gas station, police station, etc.; it could also be a bus or taxi park. In any case it was always easy to find or identify. The starting point was used to determine from where the interviewing will begin. The field manager geo-located this starting point and entered the correct information on their community tracking sheet.

Movement from the starting point:

1. Once at the given starting point, interviewers stood with their backs to the structure and moved to the right. Using the day of the interview's date code, and counting households (excluding the starting point), interviewers attempted a contact at the first house corresponding to the day of interview's date code i.e. **the date** that the interviews were being conducted determined which house was **contacted first after the starting point**.

2. This household was considered as the **first** household, and subsequent households were selected by observing the pre-determined skip.[3] It is important to keep in mind that we counted individual households and not houses, as a house/building can contain numerous individual households. However, because this study took place in rural areas, a house/building could constitute a household.

3. Interviewers kept to the right side of the road, and did not cross the street.

4. If the interview in a household selected was unsuccessful, interviewers contacted the immediate next household in his/her journey plan; filled in the reason for not being successful in the earlier house and repeated the above process till s/he was successful. Before moving on to a new household the interviewer completed a household identification sheet.

5. When an interview was completed (the interview was successful), then the interviewer observed the skip in his/her journey plan to locate the next household.

2.2 Description of the Process for Selecting Treatment Communities for the Program and the Corresponding Population of Potential Control Communities

Since the process used to select treatment communities has a strong effect in creating differences between the treatment and control populations, and it was important to take these differences into account in both the design and analysis of the evaluation, it follows that it is important to have a good understanding of this selection process. The paragraphs that follow describe the selection process.

In the water-supply program, treatment communities were selected to achieve a high impact for the program, relative to a number of factors. The process used to select treatment communities is described in the project Terms of Reference. The selection process takes into account factors that are

[3] The skip is the number of households skipped between households selected to be interviewed. It was computed as the estimated number of households (2000 population divided by the average household size in the community's zone as reported in the GLSS5+ baseline report) divided by 11, the number of interviews to be completed.

considered important to outcome, as well as previous community participation in development projects. A brief description of the selection process is as follows. The process is described in additional detail in the project Terms of Reference (pp. 8-9).

1. The Community Services Project allocated funds to agricultural communities where Farmer-Based Organizations (FBOs) are present in three geographic areas (Afram Basin Zone, Northern Agricultural Zone, and Southern Horticultural Belt). Funds were allocated to all districts within these areas based on a weighted average of the concentration of FBOs and the poverty index as measured by the Ghana Statistical Service in 2003.[4] Communities were then classified as eligible or ineligible based on whether there was at least one FBO in the community.

2. Within each district, all eligible communities were ranked according to six factors that relate to need and previous community participation in development projects. Data on these factors were obtained by MiDA teams visiting each community; these data are not available for other communities. The factors, and the maximum number of points associated with each factor, are as follows:

    a. Adequacy of water source (if water source is available). This is based on borehole or standpipe for each 300 persons, water quality, and difficulty of raising water from source to surface.
        i. Inadequate: 15 points
        ii. Adequate: 0 points
    b. Presence of guinea-worm disease (if water source is not available). This is degree based on cases reported to the health services.
        i. Guinea-worm endemic: 50 points (maximum)
        ii. Not Guinea-worm endemic: 30 points
    c. Water quality. This is based on observation, e.g., source being a pool created by blocking a stream.
        i. Not acceptable: 10 points
        ii. Acceptable: 0 points
    d. Distance to water source
        i. Above threshold: 20 points
        ii. Below threshold (<500 meters): 10 points
    e. Previous community participation. This is based on whether the community had raised its matching share for other improvements in the past.
        i. Good: 15 points
        ii. Poor: 10 points

[4] The development of the poverty indices are described in Coulombe (2005).

The maximum total number of points is 110. The communities within each district were ranked in order of point score, the cost of the water system improvement appropriate for the community was estimated, and communities selected for improvement in order of point score, starting from the highest score, until the allocated district budget was used. Note that the strict ranking was not always followed because priority was given to communities that qualified for school rehabilitation/reconstruction assistance which did not have water and sanitation facilities.

2.3 Sample Design Details

In support of the survey design effort, information about program treatment areas was obtained from MiDA program personnel and GIS data from government agencies with MiDA assistance. Using this information, NORC constructed a list of 142 treatment communities and 388 control communities, for a total of 530 communities. This list was the sample frame (sampling frame; frame) for the survey, i.e., the population from which the sample were selected. For these communities, data were assembled on the following variables (field name used in the database followed by description). The variables have been grouped into three categories: demographic/location variables; water-system-related variable; and physiographic variables. (These categories will be referred to later, in discussing matching and stratification.)

Demographic/location variables

REGION: Region of community/village
RegCode: Region code
DISTRICT: District of community/village
DistCode: District code
POVERTY: Poverty Index
COMMUNITY: name of community/village
DMS_LAT: Latitude in degrees, minutes and seconds
DEC_LAT: latitude in decimal degrees
DMS_LONG: longitude in degrees, minutes and seconds
DEC_LONG: longitude in decimal degrees

Water-system-related variables

POP_2000: 2000 Census population
NO_OF_FBOS: Number of Farmer-Based Organizations (FBOs) in community
GUINEA_WOR:[a] Guinea Worm Endemic
INADEQUACY:[a] Inadequacy of Water
WATER_QUAL:[a] Water Quality
DISTANCE:[a] Distance to water source
PREVIOUS_C:[a] Previous Community Participation
SCORE: Selection-criterion Score (based on five preceding scores)
BH: BH Intervention Type (boreholes with hand-pumps)
PE: PE Intervention Type (existing pipeline extension)
ST: ST Intervention Type (small-town water scheme)
SELECTED: Selected for improvement by MiDA if 1, not selected if 0
QUALIFIED: Eligible for improvement by MiDA
OTHER_INTE: Eligible for other MiDA-financed infrastructure improvement

a. Selection variable on local water system gathered through on-site visits described in Section 2.2.

Physiographic variables

ELEVATION: elevation in meters
SLOPE_CLASS: code for mean topographic slope of locality (codes: 1 = average slope of 0-2%; 2 = 2-4%; 3 = 4-8%; 4 = 8-16%; 5 = 16-30%; 6 = 30-45%; 7 = > 45%)
AG_SUIT: agricultural suitability index, ranges from 1-100, with 100 having highest agricultural suitability
ECOZTM_ID: econzone code (1 = Sudan Savanna Zone; 2 = Guinea Savanna Zone; 3 = Transition Zone; 4 = Deciduous Forest Zone; 5 = Moist Evergreen Zone; 6 = Coastal Savanna Zone; 7 = Wet Evergreen Zone)
NAME: Name of ecozone (ECOZTM_ID)
AGRICULTURAL GROWING PERIOD: Growing Period Pattern Zone Code - ranges from 1-3 with 3 having best growing period conditions, and likely best agricultural conditions (1 = one growing period per year 75% of years, 2 growing periods 25% of years; 2 = one growing period 55% of years, 2 periods 35% of years, 3 periods 10% of years; 3 = 2 growing periods 50% of years, one period 30% of years, 3 periods 20% of years)
THERMAL ZONE: Thermal zone: number of days per year hotter than 35 degrees C, higher the code value the hotter (code: 1 = 0 to 5 days per year; 2 = 5 to 30; 3 = 30-90; 4 = 90-150; 5 > 150)
RELATIVE HUMIDITY: Index of Annual Relative Mean Humidity: higher the index value, the more humid (can have non-linear impact on agricultural productivity, but in general the more humid the better for agriculture)
ANNUAL_RAINFALL: Average Annual Rainfall in mm per year
TEMPERATURE: Average annual temperature in Degrees Celcius
LANDCOVER: Landcover class from FAO Land Cover Dataset
FOREST: 1 if forest, 0 if not forest

Of the preceding variables, the numeric ones are potential "design variables," viz., numerical variables that may be used for matching and stratification.

The first step in the sample design process was to assess the degree of association, or correlation, among the (potential) design variables, for the treatment population. The strength of correlation among the design variables suggests how the variables should be handled in the matching and stratification process (for example, highly correlated variables may be combined, or one or more of them omitted, or a stratification determined to reduce the correlation among them (orthogonalize them)). The measure of correlation used was the Cramer coefficient of correlation, calculated for recoded variables in which each variable was coded into two categories, above and below the mean (i.e., the recoded variable has the value 0 if the value of the unit is below or equal to the mean, and the value 1 if the value of the unit is above the mean). The correlations are shown in Table B.1[5], "Basic Statistics for the Treatment Population."

The method used for matching was to form a generalized distance measure, based on all of the preceding design variables, that reflects the similarity of two population units. The distance measure is a linear combination of the normalized absolute difference between two population units on each

[5] Because of their large size, the table for this chapter are in Annex B.

variable (normalized so that the distance component for each variable varied between zero and one). The distance measure took into account the relative importance of each variable in affecting outcome, through the coefficients of the linear combination. The coefficients were a set of positive numbers (scaling factors, subjective weights). [6]

To determine the best match for a particular unit, the distance was calculated from that unit to every other unit, and the closest unit was selected as the match. This method of matching is called "nearest-neighbor" matching.

Before calculating the distances, most of the variables were recoded into sets of up to nine different values. These recoded values are used both for matching and for stratification.[7] The recoding of the variables is as follows (code value followed by original values; intervals open on left and closed on right):

RegCode: 0-2 for the three regions
DistCode: 0-22 for the 23 districts of the MiDA program areas
POVERTY: 0: 31-41; 1: 41-51; 2: 51-62; 3: 62-72; 4: 72-82; 5: 82-92.
POP_2000: 0: 0-602; 1: 602-1405; 2: 1405-2809; 3: 2809-4214; 4: 4214-7424; 5: 7424-38,927.
NO_OF_FBOS: not recoded (values 1-7)
GUINEA-WOR: 0: 0-5; 1: 5-10; 2: 10-15; 3: 15-20; 4: 20-25; 5: 25-30; 6: 30-35; 7: 35-40; 8: 40-45.
INADEQUACY: 0: 0; 1: 0-5; 2: 5-10; 3: 10-15.
WATER_QUAL: 0: 0; 1: 0-5; 2: 5-10; 3: 10-15; 4: 15-20.
DISTANCE: 0: 0-5; 1: 5-10; 2: 10-15.
PREVIOUS_C: 0: 0; 1: 0-5; 2: 5-10; 3: 10-15.
SCORE: 0: 0-50; 1: 50-55; 2: 55-60; 3: 60-65; 4: 65-70; 5: 70-75; 6: 76-80; 7: 80-85; 8: 85-90; 9: 90-95.
BHcode: not recoded (values 0-1)
PEcode: not recoded (values 0-1)
STcode: not recoded (values 0-1)
OtherIntCode: not recoded (values 0-1)
ELEVATION: 0: 77-365; 1: 365-653; 2: 653-940; 3: 940-1228; 4: 1228-1515; 5: 1515-1802.
SLOPE_CLASS: not recoded.
AG_SUIT: 0: -12-4.5; 1: 4.5-21; 2: 21-37.5; 3: 37.5-54; 4: 54-70.5; 5: 70.5-87.
ECOZTM_ID: not recoded.
AG_GROWING_PERIOD: not recoded.
THERMAL_ZONE: not recoded.
RELATIVE_HUMIDITY: 0: 23-33; 1: 33-43; 2: 43-54; 3: 54-64; 4: 64-74; 5: 74-84.
Precip (PRECIPITATION): 0: 900-1000; 1: 1000-1100; 2: 1100-1200; 3: 1200-1300; 4: 1300-1400; 5: 1400-1500; 6: 1500-1600; 7: 1600-1700.
TEMPERATURE: 0: 24; 1: 25; 2: 26; 3: 27; 4: 28.
LANDCOVER: 0: 14-47; 1: 47-79; 2: 79-112; 3: 112-145; 4: 145-177; 5: 177-210.
FOREST: not recoded.

[6] The importance weights may be specified in any scale – before being used to calculate distances, they are normalized to sum to one.
[7] Recoding is not necessary for matching, but it generally simplifies the design process.

Recoding was done only for variables on the interval scale of measurement. The variable DistCode was not recoded. It has 22 different values (0-22, for the 29 districts of the MiDA program area), and is nonordinal (nominal). It was used only in matching, not in stratification. For most variables on the interval scale of measurement, the category boundaries used in recoding were determined simply by dividing the range of the variable into a small number of categories of equal width. This approach is useful when the data are to be used to support development of a parametric model, such as by using multiple regression analysis. If this procedure is seen to result in an inordinate number of units in a particular category, the category boundaries would be modified as appropriate.

Data were missing for a few of the variables, and missing values were imputed either by using the median of the non-missing values, or the median of the observations in the same ecozone (same Ecoztm_Id code), if that value was present.

For stratification, two competing objectives came into play. On the one hand, it was desired to have a modest amount of variation in each design variable, to enhance the precision of estimates of the relationship of outcomes to the variable. On the other hand, it was desired to keep the probabilities of selection of households as uniform as possible, to achieve high precision for unbiased (weighted) estimates of impact (at the household level).[8] Both of these competing objectives were accommodated by the following procedure. The expected allocation that would result from simple random sampling of primary sampling units was determined. If this allocation had inadequate variation, it was adjusted by increasing the probabilities of selection for strata containing few units and decreasing the probabilities of selection for strata containing many units. Second, the stratum allocation for POP_2000 (which is approximately proportional to the number of households in a community) was determined in a way that approximates sampling of communities with probabilities proportion to size (population). Since a fixed number of households is selected from each sample community, this results in household probabilities of selection that are approximately uniform. Because outcomes of interest are expected to relate to community size, it was desired that the allocation of sampling effort to the size strata be somewhat balanced, and so the allocation was modified somewhat from the proportional allocation.

For most of the design variables, the variation that would result from simple random sampling was considered adequate. It was decided that stratification was desirable for only two variables: POP_2000 and SCORE (which is considered a good indicator of outcome).

Some additional information will be provided on the procedure used to determine the stratum allocation for POP_2000. Since the unit of analysis is the household (not the community), it was

[8] In the unbiased estimates, the sample values are weighted by the reciprocals of the probabilities of selection. If the weights vary tremendously, the standard error of the estimate is large.

desired that the selection probabilities for households be as uniform as possible, subject to variations required to achieve stratification objectives. In a descriptive survey, uniformity in the household selection probabilities would typically be achieved (in a design such as the present one, which selects a fixed number of households from each sample community) by selecting communities with probabilities proportional to size (where the measure of size was community household population, i.e., number of households). That approach does not work here, since the selection probabilities are fixed to achieve expected marginal stratification constraints at the community level, not to make the probabilities of selection uniform at the household level. This latter objective is achieved instead by *stratifying* on the number of households (or its surrogate, POP_2000). A good method for setting the stratum boundaries for an analytical survey design in which the household is the unit of analysis is to set the stratum boundaries such that there are approximately equal numbers of households in each stratum. This method was used for this application. As mentioned, we depart somewhat from the proportional allocation to achieve better "balance" of the strata.[9]

For stratification and matching, it is not necessary to use a large number of categories. Experience has shown that having about three to six categories is usually sufficient. Using a larger number of categories simply adds complexity without achieving additional improvement in reducing bias or increasing precision. The use of a larger number of categories (e.g., six in the present case, versus, say, three) enables a higher degree of precision in matching, and a higher degree of control of sample size for categories of interest.

For each treatment unit, a probability of selection is determined such that every treatment unit has a nonzero probability of selection and such that there is a certain amount of variation in each variable. In this application, there was no need to attempt to orthogonalize the design variables. The sample of treatment units is selected by generating a set of uniformly distributed (pseudo) random variable, one for each treatment population unit, and including the unit in the sample if the corresponding random number is less than the probability of selection. (To facilitate the making of replacements, the population is put in random order prior to the sample selection.)

To facilitate matching, the design variables were classified into three groups of variables related to each other from the viewpoint of a causal model relating them to outcome, or to administrative convenience (e.g., efficiency in sampling). These three groups are the groups specified above in the listing of the design variables, viz., Group 1: demographic/location variables; Group 2: water-system-related variables; and Group 3: physiographic variables. Group 1 was assigned a relative weight of 100; Group 2 a relative weight of 100; and Group 3 a relative weight of 50. Within these three groups, the relative weight of each component variable is as specified in Table B.2, "Matching

[9] Because outcome measures of interest are expected to be related to community size, it is desirable to have a certain amount of variation in the variable, to increase the precision of the estimated relationship.

Importance Weights." The group weight is split among the members of the group. Within a group (except for Group 1, which relates to administrative convenience), variables (or subgroups of related variables) considered to have the strongest relationship to outcome are assigned the highest weights. It is noted that the results of matching were not highly sensitive to modest variations in the relative magnitudes of the weights, especially if the ranking of the weights remains the same.

This matching process used is called "greedy matching." With greedy matching, at each stage of the matching process the best possible (nearest-neighbor) match is sought for a unit, out of the population of all unmatched units. This differs from "optimal" matching, in which it is attempted to find a set of matches that are good in some overall sense. Before starting the matching process, as mentioned earlier, all population units were arranged in random order, and a random sample of treatment units was selected in accordance with the variable probabilities of selection of the design. The set of match-mates was selected for each treatment unit, starting at the top of the list, by examining all other units and selecting the one that was closest (with respect to the generalized distance measure). Since greedy matching is used from the top of the list, this process assures that at each step, the best possible match was made from the remaining, as-yet-unmatched units. This process is preferable to "optimal" matching.[10]

Annex Tables B.3-B.14 provide details on the sample design and selection process.

Table B.3, "Population Frequencies (Treatment Units)," shows the frequency distribution of the treatment-community population over the stratum cells defined by the coding scheme specified above. Table B.4, "Population Frequencies (Control Units)," shows the frequency distribution of the control-community population over the stratum cells. The "supports" of the two distributions are rather similar, which indicates that it should be possible to do reasonable job of matching.

It is of interest to consider what sort of sample would result if simple random sampling were used to select the household units, or probability-proportional-to-size (pps) sampling were used to select the treatment community sample, instead of a stratified ample. From Table B.3, we see that there is actually a reasonable degree of spread in all design variables, and this would have been reflected in a simple random sample. The distribution of the population over the design variables is not at all uniform, however, so that a simple random sample would not exhibit good balance over the design variables. Table B.5, "Expected Sample Frequencies for Simple Random Sampling of Households or Probability-Proportional-to-Size Sampling of Communities (Treatment Units)," shows the expected sample frequencies corresponding to simple random sampling (or pps sampling).

[10] Note that this procedure (greedy nearest-neighbor matching) is not "one-to-one exact" matching, since with the small populations involved, it is not possible to find many treatment/control pairs that match exactly on all match variables, even when recoded into a small number of categories.

Table B.6, "Desired Sample Frequencies before Matching (Treatment Units)," shows the desired allocation of the treatment sample units to the various stratum cells, for the marginal stratification defined by each design variable. This allocation allocates the sample to the stratum cells of the design as discussed earlier. There is a reasonable amount of variation (if not so much balance) in most of the design variables, so that control of variation by stratification is not necessary. The only two variables selected for stratification are POP_2000 and SCORE.

Table B.7, "Desired Sampling Fractions before Matching (Treatment Units)," shows the minimum desired proportion of treatment units (sampling fraction) in each stratum cell to be included in the sample. For variables not subject to stratification, all of the sampling fractions are zero. This table shows that the sampling fractions vary considerably from uniform (a constant), in order to achieve the stratification objectives of the design. While achievement of uniform selection probabilities is often an objective for descriptive surveys, it is of secondary importance for an analytical survey – achieving a high level of precision for the double-difference estimate of impact, or a high level of power for testing hypotheses about the double difference (such as detecting a double difference of a specified size) usually does not correspond to uniform selection probabilities.

Once Table B.6 has been specified, the probabilities of selection were determined for each treatment unit of the population, to achieve, as closely as possible, the desired stratification. Table B.8, "Expected Sample Frequencies before Matching (Treatment Units)," shows the expected number of treatment sample units in each stratum cell, corresponding to these selection probabilities. A random sample of treatment units was then selected, using the determined probabilities of selection. Table B.9, "Actual Sample Frequencies before Matching (Treatment Units)," shows the distribution of the selected sample (of treatment units) over the various stratum cells. It is not exactly equal to the "desired" allocation, but it has substantial spread in all design variables, which is the essential requirement. It is much more balanced than the general population.[11]

For each treatment unit selected for the sample, a matched control unit was selected, according to "closeness" as measured by the generalized distance measure described earlier (i.e., the nearest neighbor is selected). As discussed earlier (and shown in Table B.2), all design variables are involved in the matching process. The resulting distribution of treatment sample communities over the stratum cells is shown in Table B.10, "Actual Sample Frequencies after Matching (Treatment Units)." Table B.11, "Actual Sample Frequencies after Matching (Control Units)," shows the distribution of control sample communities over the strata. These two distributions are similar because the matching was effective. These two distributions are much more similar than the original distributions of the

[11] It differs somewhat from the desired allocation both because of "sampling fluctuations" and (more significantly) because it is not possible to find population units that have the combinations of match variables required to satisfy the marginal-stratification requirements.

populations of treatment and control units (shown in Tables 3 and 4). The similarity of these distributions is evidence that the matching procedure may be effective in reducing selection bias associated with variables known prior to the survey.

Table B.12, "Table Showing Matched-Pair Details," presents detailed information to show the quality of the match. This table summarizes how close the control community is to the treatment community, for each match variable. The sample units are sorted into match sets (i.e., are sorted by MatchSetNumber). The last columns of the table are the coded values of the match variables. The table shows that the matched pairs often match on the values of the coded match variables. The match is better for the match variables having high importance weights. With the large number of match variables used in the present application, it is not expected that there would be a high proportion of exact matches.

Map 2.1 shows the locations of treatment and control communities.

Table B.13, "Match Quality Summary," summarizes the overall quality of the match by showing the average distribution of distances between the treatment sample community and its matched control community, for each coded match variable. This summary shows that there is a substantial proportion of exact matches. As noted, for an application such as the present one, in which there are many match variables, it is not expected that there would be a very high proportion of exact matches. (The distance is meaningful only for ordinal variables, not for nominal variables (such as DistCode).) Note, however, that many matched pairs have exactly the same value of DistCode.

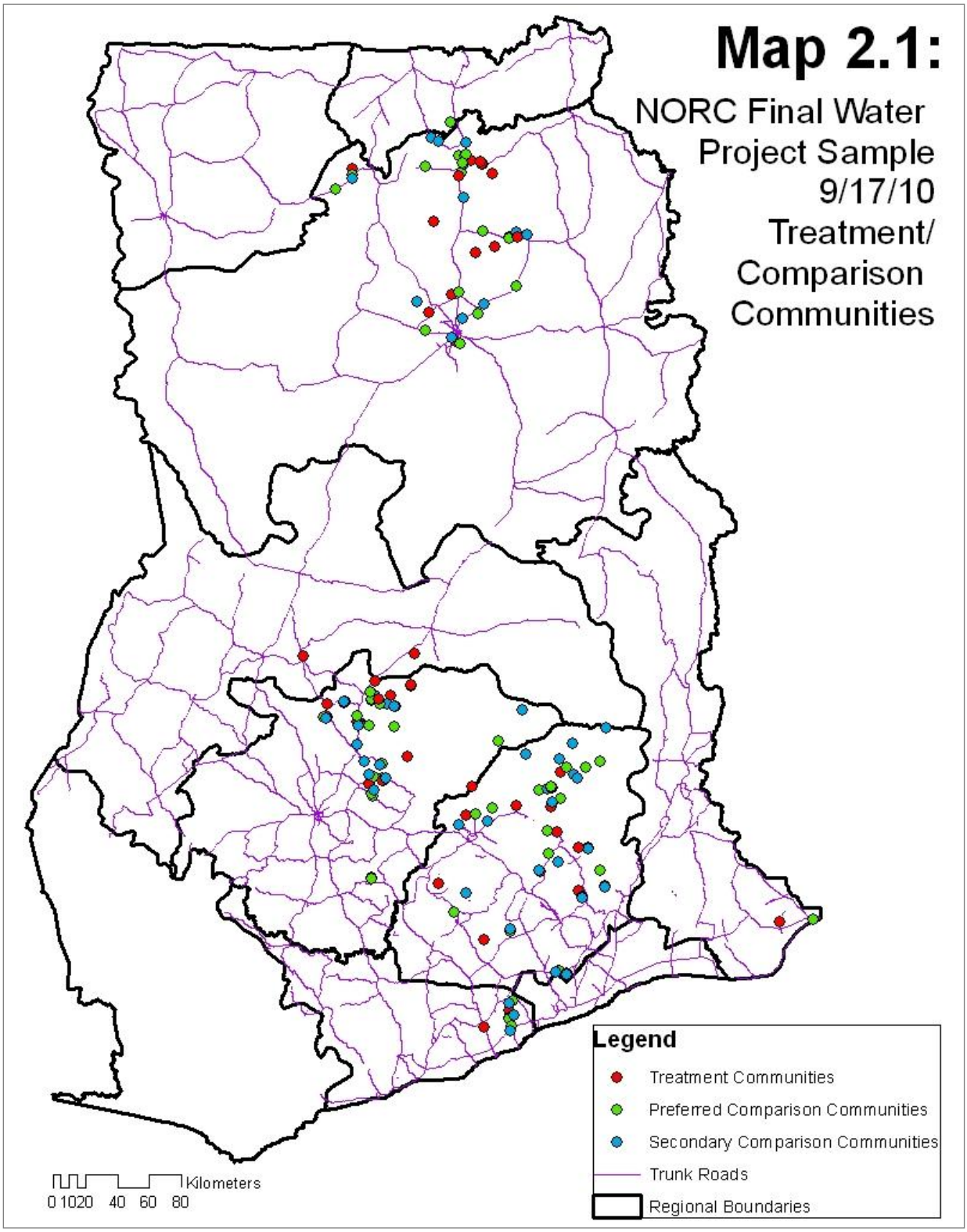
Map 2.1:
NORC Final Water Project Sample
9/17/10
Treatment/ Comparison Communities
Legend
Treatment Communities
Preferred Comparison Communities
Secondary Comparison Communities
Trunk Roads
Regional Boundaries
Kilometers
0 10 20 40 60 80

Table B.14, "Community Sample for the MiDA Water Supply Activity Survey," presents a list of the treatment/control sample triplets selected for the sample. Each match set consists of a treatment unit and two matching control units. The table contains 180 communities, sorted into 60 match sets, which are identified by the MatchSetNumber (which is the IDNO of the treatment unit). The desired sample of 50 match sets is selected as the first 50 match sets (in order of MatchSetNumber) that are acceptable (i.e., as discussed earlier in the paragraph about treatment of nonresponse, for which the treatment unit is acceptable). The column "InSample" specifies the nature of the sample units in each matched pair: 1 indicates a treatment unit; 2 indicates the first-selected control unit; and 3 indicates the second-selected control unit.

3. Questionnaire Development

The task of the questionnaire is to obtain the information necessary to test the hypotheses. In general in developing a questionnaire one tries to employ questions whose reliability has been proven in earlier surveys. In this case the TOR instructed us to the extent possible to use the questions from the fifth round of the Ghana Living Standard Survey as adapted for use in gathering baseline information in the areas in which MiDA is operating for the overall evaluation of MiDA's programs. Using these questions for the present evaluation permits tighter comparisons from the baseline data and this evaluation's data than would otherwise be the case. This survey is generally called the GLSS5+. We have taken questions from the GLSS5+ from the household survey and from the community survey. (The questionnaires are in Annex A.)

We also drew on three other sources in developing the questionnaire, particularly the sections inquiring about households' use of water and the consequences of greater or lesser quantities of water being available and the way the water is used.

First, we consulted the questionnaire employed in the baseline data collection for the impact evaluation of the Millennium Challenge Corporation (MCC)-supported water improvement project in Lesotho.

Second, we studied the structure and the precise wording of questions employed in water usage surveys conducted by UNICEF, the Development and Health Survey, and the Environmental Health Project (USAID) that have set international standards and conventions.

Third, we employed some questions from the Core Welfare Indicators Questionnaire-II (CWIQ). As indicated in Chapter 1, one impact that the evaluation should assess is that on household well being, which was taken here to be household consumption expenditures. Analysts face the conundrum of measuring consumption expenditures when the interview time available to explore this topic was sharply limited. In our case, the total interview time budgeted was 90 minutes of which the majority is allocated to the all-important water sector questions.

After consultation with the Ghana Statistical Service (GSS) and experts at the Institute of Statistical Social and Economic Research (ISSER), it was concluded that the most effective approach would be to use an already estimated regression model that relates consumption expenditures to independent variables measuring household well-being, such as ownership of durables and the size of expenditures on certain goods.

In particular, such a model was estimated using GLSS4 survey data that included the variables used in the much smaller CWIQ. The resulting regression model, estimated in 2005, has 13 independent variables (GSS, 2005, xxi-xxiv). Analysis of the model's predictive ability as reported by GSS was encouraging.[12] Because this is a difference-in-differences analysis, there is no need to adjust the expenditure values for inflation since the model was estimated. It is the real difference between the estimated baseline and follow-up values that is of interest.

ISSER also reported in a meeting in August, 2011 that staff was in the process of estimating a similar model with GLSS5+ data. If this new model is available on time for the impact analysis and it uses variables for which this project gathers data, we intend to employ it in the analysis.[13] Given the strategy just outlined, the survey includes questions to obtain information to construct the variables in the model using data from the CWIQ.

The regression model was structured to estimate household consumption at a point in time. A critical question was whether there would be sufficient changes in the independent variables between the baseline and after-water improvement survey to provide an accurate estimate of the change over a period as short as 18-24 months. Several variables measure the quality of housing amenities durable goods, e.g., the type of toilet facility, use of electricity, ownership of an electric iron, refrigerator, television, etc., and another bloc focuses on expenditures for necessities per adult equivalent, e.g., rice, bread, kenke, tomatoes, and soap. It appears that these variables should be quite sensitive to changes in economic resources and therefore provide a reasonable estimate of change over the observation period.

Our questionnaire also includes some questions on the household's income obtained by asking about the income of each employed person in the household. Space was insufficient to inquire about various classes of consumption expenditures.

[12] Similar results are reported for similar analysis also for Ghana by Fofack (2000). Fofach's work is cited by GSS as being the basis for the model presented in its publication. Goodness of fit information, i.e., for the observed and imputed values being in the same (main diagonal) or adjacent diagonals for expenditure quintiles presented in the GSS report is as follows:

Geographic area	Main diagonal (perfect match)	Adjacent diagonals (near perfect match)	Total
National	50.5	39.7	90.2
Urban/rural	51.6	39.5	91.1
Strata (7)	49.3	39.4	88.7
Region (10)	44.4	41.8	86.2

[13] ISSER staff said that they would alert us when the analyst's work is quite advanced and it may be possible for them to estimate a version of the model using variables included in our baseline survey.

Questions on water supply, cost and usage differ in several areas from those in the GLSS5+ in order to obtain more accurate and comprehensive information on time spent acquiring water, the quantity acquired, the source of water by the purpose for which it was used, and hygiene related to water usage. Some additional questions on household hygiene areas were also included so that in the statistical analysis of change the role of increased water supply per se in effecting health outcomes can be disentangled from hygienic practices.

The survey was being conducted during Ghana's rainy season. This is a problem because the use of various water sources will be different from the dry season. With households using collected rain water, for example, the impact of the program will be underestimated if we ask only about current water sources and quantities. The dry season is November-February. The survey could not be delayed until this time because the drilling of boreholes and other improvements were already underway, i.e., we would have lost the "before" observation on treatments if the survey were delayed. Despite the probable problems with respondent recall, it was decided to ask about dry season water sources and water use patterns in the survey so as to obtain a more accurate estimate of the benefits flowing from the investment program.

The necessity of documenting the change in dry season water consumption patterns for the evaluation argues for conducting the second ("after") survey round in the dry season. It is probable that it will take households a very short time to adjust their water acquisition patterns to use the improved water sources.

4. Data Collection

4.1 Development of Operational Plan and Manuals

NORC prepared a detailed operational plan for data collection and data delivery. The operational plan included the following:

- Roles and responsibilities of field staff. NORC developed detailed position descriptions for all field workers. The position description included identifying the project objectives, defining the scope of the project, listing all critical project deadlines, stating the client and project stakeholders, listing the key roles and their responsibilities, creating an organizational structure for the project, documenting the overall implementation plan, and listing any risks, issues and assumptions that are necessary for the field workers to understand in order to carry out their duties at the highest level.
- Data security plan. According to the International Compilation of Human Research Protections, 2008 Edition, compiled by the Office for Human Research Protections of the U.S. Department of Health and Human Services, Ghana does not have any statutory requirements for security of sensitive survey data. The operational plan therefore developed a data security plan that met standards for U.S. institutional review boards with respect to data sets containing sensitive data and management of paper forms containing sensitive data.
- Field work schedule and management reporting. NORC developed a schedule for each field team and outline the case management process for tracking the status of cases (complete, incomplete, refusals, unavailable), the need for any follow-up interviewing, and reporting on data collection progress.
- Defining quality assurance. NORC defined a plan for reviewing data as it was collected. This plan included daily reviews by field supervisors and coders and frequent reviews of data extracts by NORC and Panafields staff as well as case validation.
- Data documentation and delivery. NORC developed a detailed data documentation plan following MCC guidelines outlining the modules needed for final documentation and the steps involved in producing those modules before final delivery.

NORC built on its extensive experience in both developing training materials and implementing field training in international contexts to ensure the process ran efficiently. NORC's survey staff worked closely with the Panafields team on development of manuals and will be in Ghana for the training, pre-test, and first week of data collection to provide support to Panafields. Interviewer and supervisor manuals address all aspects of data collection, beginning with identification of eligible households,

how to implement a systematic random sampling plan, data collection once identification is completed, and all aspects of questionnaire administration and human subjects protections. The interviewer manuals include modules on gaining cooperation, interviewing techniques, a study overview, section by section description of the questionnaire, keeping records and bookkeeping, respondent confidentiality, and exercises to utilize each job skill learned. The importance of recording clear and concise information was stressed during the training. Naming protocols and accurate data entry were also stressed.

4.2 Institutional Review Board (IRB)

An IRB is a group of senior survey experts that reviews all survey plans to ensure that respondents' privacy is respected and that access to the collected data is such that individual respondents cannot be identified.

NORC has extensive experience in preparing the documentation necessary to secure Institutional Review Board (IRB) approval. NORC has its own IRB, which follows a formal process for examining all research projects to assure human subjects protection. NORC's IRB is registered with the HHS Office of Human Research Protection and has a Federal-wide assurance (FWA 00000142). The NORC IRB Administrator and Chair are responsive to the need for timely reviews, and all Board members take an active role in helping guide protocols to meet the highest standards for human subject protections. NORC's IRB requires that research protocols provide sufficient detail to ensure that (1) the selection of subjects is equitable, subjects' privacy is protected, and data confidentiality is maintained; (2) informed consent is written in language that study participants can understand and is obtained without coercion or undue influence; and (3) appropriate safeguards protect the rights and welfare of vulnerable subjects. Prior to beginning field work, NORC submitted a formal research protocol that provided the purpose of the evaluation, procedures to which respondents are subjected, and the research benefits and risks. The protocol detailed the methods used to ensure confidentiality of the data and the process of obtaining informed consent from respondents.

4.3 Recruitment and Training of Field Staff

The key to the successful execution of a survey is in the quality, commitment and training of the field staff—field interviewers and supervisors. Field interviewers must be drilled to deliver the questions in exactly the way that they were designed and must fully understand the meaning and context of the questions. The uniformity of survey application is best ensured by keeping the field team as small as possible consistent with the time available for the study.

Panafields, with the oversight of NORC, was responsible for advertising, interviewing, and hiring all interview and data entry team members. To the maximum extent possible, Panafields drew on its

roster of field interviewers and supervisors with whom it has previously worked in order to ensure the highest level of field staff quality. We recruited the appropriate number of team leaders and interviewers to oversee and conduct the survey tasks within a relatively short three-week data collection period. All field interviewer and supervisor candidates were interviewed by Panafields' Data Collection Field Manager to establish their experience, interpersonal skills, understanding of the basic concepts used in household surveys, ability to record accurately information on the questionnaires, capability to identify the appropriate people for the interviews, professionalism and neutrality, and capacity to understand the necessity of avoiding directing the respondents replies. Interview staff had to be proficient in the local language of the areas in which they will be conducting the survey.

For the Supervisor candidates, their leadership qualities and objectivity were evaluated. Supervisors were responsible for ensuring that respondents are correctly identified, making certain interviewers comply with all consent and confidentiality requirements as approved by the IRB, and to verifying the completeness and internal consistency of the questionnaires before they are returned from the field to the central office for data entry.

Based on these criteria, the best candidates were invited to participate in the training. More field interviewers than required were invited for the data collection to attend the training. The purpose of these supplemental interviewers was to allow for possible attrition during the training process as well as to supply replacements in the event field interviewers drop out during the data collection period.

We employed six field teams consisting of four interviewers plus one supervisor and one editor. The supervisor and editor were in charge of ensuring that the systematic random sampling approach was applied correctly and that identified households were eligible to be interviewed. After identifying an eligible household, enumerators were sent to the household to conduct the actual interview. The households were marked with a geo location after accepting to take part in the survey and field mangers located each household for the interviewers.

We estimated that each interview would take approximately 2 hours (including time for locating respondents and gaining cooperation). With a sample size of 100 communities (50 treatments and 50 controls) and 11 completes needed per community per day we expect each team to complete one community per day.

Training the field teams was a critical component for ensuring data quality. A successful training provided the interviewer with a clear understanding not only of the design and content of the instruments, but also on how to administer the instruments in a manner that avoids introducing bias into the responses.

NORC conducted a two stage training whereby team leads are trained first by an experienced NORC field staff trainer (Training of Trainers, or TOT). In the first stage Team leaders received comprehensive field management training from NORC's survey expert on all aspects of the project and data collection tasks, including enumeration, sampling, case management, quality control, field supervision, and interviewer training. This training took one day and gave us a pool of individuals who were then tapped to take part in the training of field staff.

The second stage of training required four days and both interviewers and supervisors participated (Table 5.1). We initially planned for three days of training but scheduled an additional day to increase enumerator confidence and quality. The training began with an introduction to the Water Survey and the goals of the project. It was important that interviewers as well as supervisors understood and believed in the study in order to ensure the highest level of data collection quality. In training, we needed gain the interviewers' cooperation and interest from the start. Over the course of training, interviewers learned how to gain cooperation; determine what constitutes an eligible unit for sampling (both for the household and community surveys), administer the survey to the appropriate respondents; correctly enter survey information; and conduct record keeping of all visits and contacts with the sampled respondents.

Table 4.1: Training Schedule

Day	Agenda
Day 1 – Field Managers	• Opening and logistics of training • Introduction to Water Evaluation Survey • Survey design and methodology • Field sampling and enumeration • Item-by-item review of the Water Survey • Identifying respondents for the Community Survey • Item-by-item review of the Community Survey • Field supervision and Field Editing • Tracking and Reporting
Day 2 – Interviewers & Managers	• Introduction to Water Survey • Survey design and methodology • The Household Characteristics Module • The Education Module • The Health Module • The Time Use and SES Module • The Facilities and Water Module • Role playing and interview techniques • Sampling exercises
Day 3 – Interviewers & Managers	• Interviewing Techniques • Survey Logistics • Survey Review • Exercises and role playing
Day 4 – Interviewers & Managers	• Interviewing Techniques • Survey Logistics • Survey Review • Exercises and role playing
Day 5 – Interviewers & Managers	• Field Pilot Test
Day 6 – Interviewers & Managers	• Field Pilot Test • Pilot Test Debriefing • Interviewer Assignments

NORC's Interviewer training emphasized using the skills field staff are learning. The training required that interviewers perform repeated exercises to ensure that they fully understand the materials and are able to apply what they have learned in the field. Developing modules that test and hone the skills interviewers need in the field was a key component of NORC's training plan. Field Staff were required to succeed when performing these exercises or have to repeat them until they are able to implement the skills at the highest possible level.

Following the classroom training, there were two days for a pilot trial to give enumerators the opportunity to test the survey instruments and protocols under realistic conditions (and adjust thereafter as needed). During the pilot each field interviewer and supervisor were observed by the

NORC/Panafields team responsible for managing the field work. The pilot took place in communities near Accra that were not part of the sample, but that had similar characteristics. When the pilot was concluded, the best candidates were chosen as field interviewers/supervisors and substitutes identified.

4.4 Implementing Data Collection

The field teams were supplied with their assignments and schedules at the start of each day. All equipment was provided to the field teams at this time as well. Each field team was supplied with enough materials (questionnaires, interview aids, etc.) to cover their daily assignment. Teams organized so that interviewers have command of the local language in each community. Team supervisors had a cell phone to be able to the field manager to solve any doubt or unexpected situation in the field.

The field interviewers collected the data and returned the completed questionnaires to the supervisor and editor for review. The editor checked the work to ensure that the proper person was interviewed and for completeness before receipting the questionnaire as a complete case. Upon the successful completion of each interview, the team supervisor identified the next respondent/location for survey administration. Team supervisors were responsible for handling issues as they arose, such as gaining cooperation, and resolving questions about survey administration.

In addition, the field manager was responsible for administering the community survey which was also edited by the field editor. Both mangers and editors were trained on administering the community survey in case of problems in the field that may limit the managers' ability to carry out the community survey.

Although we initially estimated that it would take 2 hours to administer the surveys, the actual average administration time was closer to 70 minutes. The reduced administration time most likely points to the experience of the enumerators and ease of administration.

Data collection began on September 23rd, 2010 and ended on October 10th, 2010 for a 16 day data collection period.

4.5 Supervision and Quality Control

NORC and Panafields place a strong emphasis on guaranteeing the quality of data gathered in the field. To this end, fieldwork was continually supervised by the Field Manager. Errors arising from field interviewers' mistakes are an under-rated source of error in statistical surveys. Often, great trouble is taken to design sample sizes that will produce acceptable expected errors (given expected variances) but little attention is paid to avoiding mistakes in the data gathering process as such.

During each day of data collection, team leaders collected the completed surveys from the field editors after they have reviewed each completed questionnaire. The team leader then selected a small subsample of responses (5%) for data quality follow-up. He/she revisited the households the interviewers visited to ensure that the items were collected correctly. In addition, the team leader reviewed all of the work of the interviewers each day to ensure legibility and consistency in filling out the survey forms. After field editing the surveys, team leaders and/or editors sealed the surveys in manila envelopes with the locality information filled out on the front of the envelop for delivery to the central office.

Field control sheets were updated to reflect the day's work and this information relayed to the central office at least twice a week by cell phone communication. In addition, each team was visited at least once during data collection by the central office data collection manger who collected completed forms (manila envelopes) to be returned to the central office. These forms were then re-checked by central office staff to ensure data quality.

At the end of the field period all surveys were returned to the central office in manila envelopes for data entry. Locating sheets with household identifying information are stored in the central office in a locked file cabinet between data collection phases.

Several communities needed to be replaced during the data collection period. These replacements were due to inclement weather encountered during the rainy season except in the case of Dungu and Yong. Dungu was replaced because it was no longer receiving the MiDA intervention, Yong was its associated control community. The communities that needed to be replaced were:

Santa (treatment)

Tuvuu (control)

Sodzikope (treatment)

Akroso (control)

Dungu (treatment)

Yong (control)

These villages were replaced with the following treatment and control pairs:

Nyanyano (treatment)

Gomoa Lome (control)

Tong (treatment)

Kpalguma (control)

Apaah (treatment)

Adidwan (control)

4.6 Data Processing

Data processing took place in Panafields central office in Accra. A database was created for analysis in SPSS format. The data entry system included controls designed to avoid the input of inconsistent data. Data processing consisted of the coding and input of data and in cleaning the database. To avoid the transcription errors all data were input twice by separate data input clerks and inconsistencies in data entered identified and resolved. Data cleaning was carried out for detecting and correcting, removing, or flagging incorrect data, errors in format, incomplete data, inconsistent data, etc.

After data entry quality control procedures uncovered two issues with the data that warranted further investigation. The first issue was identified as a mistake in the data entry program and was corrected when discovered. The second issue was a misunderstanding by enumerators on how to code 'legitimate skips' in the questionnaire in reference to multiple household members fetching water from one source. This issue was resolved during data cleaning. Both issues were correctable and neither issue impacted the final quality of the dataset.

4.7 Results

Survey results are summarized in Table 4.2. As shown the overall response rate was very high at 92 percent. Actual refusals were extremely rare. Overall, very strong results.

Table 4.2: Household Survey Results

Disposition Code	Description	Treatment Communities		Control Communities		All Communities	
		n	%	n	%	n	%
1	Completed	600	91.9	600	91.3	1200	91.6
2	No household member at home or no competent respondent at home at time of interview	37	5.7	35	5.3	72	5.5
3	Entire household absent for extended period	6	0.9	12	1.8	18	1.4
4	Postponed/rescheduled (interview was postponed and a new time scheduled)	1	0.2	2	0.3	3	0.2
5	Final refusal (interview was refused/no interview completed)	4	0.6	7	1.1	11	0.8
6	Dwelling vacant or address not a dwelling	3	0.5	1	0.2	4	0.3
9	Other non-interview	0	0.0	0	0.0	0	0.0
10	Partial complete/will return (interview was stopped but will continue later)	1	0.2	0	0.0	1	0.1
11	Partial complete/interview finished (interview was stopped and will not continue)	0	0.0	0	0.0	0	0.0
12	Temporary refusal (interview was refused, FS will follow-up)	1	0.2	0	0.0	1	0.1
13	Out of scope (the household is not within the sample)	0	0.0	0	0.0	0	0.0
TOTAL		**653**	**100**	**657**	**100**	**1310**	**100**

5. Baseline Data and Analysis

Introduction

The information in the tables presented here comes from two surveys described above that were conducted in September-October 2010 in 50 communities scheduled to receive improved water services under the MiDA Investment Program and 50 similar control communities. Both final questionnaires are provided in Annex A. A few questions were adjusted from those in the Inception Report during the training of interviewers and field test testing and therefore are differ from those in the questionnaires presented in the Inception Report.

Five points about the tables should be noted.

First, the tabulations are designed to give an overview of the communities and their occupants with particular attention to access to water, its quality, price, and quantity consumed. They also are designed to shed light on the baseline situation with respect to the hypotheses that the evaluation is addressing that are outlined in Chapter 1.

Second, the project's Terms of Reference emphasize the importance of comparisons between the results of the GLSS5+ data and the data from the household survey. Many, but certainly not all, questions included in the water investment surveys are the same as those in the GLSS5+ so that on this basis many results should be similar. The principal differences are in the questions asked about water sources, usage, and the like. An important difference between the surveys conducted for this analysis and the GLSS5+ is that the GLSS5+ is designed to be a representative survey of the whole area in which MiDA is working while the sample of households in the communities in the water survey only represent the population of communities receiving treatment and control communities are simply a sample of non served communities.

Most of the tables presented were formulated to correspond as closely as possible with those in the published GLSS5+ report. Where possible we employ information for households surveyed in the GLSS5+ in rural areas, but some GLSS5+ results tables do not provide such a breakout; and data are presented for all households. In general, we have aligned the data from the surveys as closely as possible, but the reader is cautioned that differences in the results from the two surveys may result from differences in the samples, the geographic basis on which they are reported, and, in some cases, the questions asked. Where questions differ substantially between the two surveys, data from GLSS5+ are not included in the tables.

Third, the terms of reference indicated that data should be provided for individual communities. This has not been done because with sample sizes of only 12 such information would have little meaning.

Fourth, results presented from our surveys are based on unweighted data.

Lastly, for some tabulations of water and sanitation facilities we aggregate facilities into "improved" and "unimproved" groups to simplify presentations. The definitions, based on WHO and UNICEF guidelines, are shown in Table 5.1.

Table 5.1: Definitions of Improved and Unimproved Water and Sanitation Facilities

Water categories	Sanitation categories
Improved sources of drinking-water:	**Improved sanitation:**
Household connections	Flush toilet
Public standpipes	Piped sewer system
Boreholes	Septic tank
Protected dug wells	Flush/pour flush to pit latrine
Protected springs	Ventilated improved pit latrine (VIP)
Rainwater	Pit latrine with slab
	Composting toilet
"Not improved" sources of drinking water	**Unimproved sanitation:**
Unprotected dug wells	Flush/pour flush to elsewhere
Unprotected springs	Pit latrine without slab
Vendor provided water	Bucket
Bottled water	Hanging toilet or hanging latrine
Tanker	No facilities or bush or field
	Public facility

In the balance of this chapter, tables based on household data are presented first and then those based on data from the community survey.

Information from the Household Survey

The first few tables provide basic information on demographics and educational attainment of those living in the sample communities. The basic breakdown of the population by sex is presented in Table 5.2, with females constituting a rather larger portion of the population than males. The pattern is the same in both data sets, although women are somewhat more dominant in the treatment and control (T-C) communities.

Table 5.3 provides tabulations on household composition, organized by the number and sex of adults in the household. For the traditional household with at least one male and one female present, about 75 percent of GLSS5+ households have children which compares with figures of over 85 percent for both groups of T-C households. It is hard to make much of this difference or others because of the differences in the geographic coverage of the two surveys. Broad patterns are similar in both.

Table 5.2: Average age of household head, by sex and locality (%)

Sex	From GLSS5+ (Rural Localities)	From HH Survey (All HH's)	
		Control	Treatment
Male	44.3	47.4	47.4
Female	51.9	53.8	57.7
Total	**46.7**	**48.9**	**50.1**

From Table 5.4 one sees that nearly one-third of males and around 45 percent of females over 15 years old in the household survey have not attended school. Only about 6 percent of those over 15 completed secondary school, with another 25 percent having achieved one of the school leaving certificates. Education achievement is broadly greater among those in the overall MiDA program area, certainly in part because it includes urban as well as rural areas.

The household survey made careful inquiries about sickness in the household during the two weeks before the interview took place—a recall period widely recognized as quite accurate. The overall (total) incidence of illness or injury reported in the two surveys is quite similar, at 13 percent for GLSS5+ and about 15 percent for the household survey (Table 5.5). Of particular interest is the incidence of illness for children under 5 years old because of the high mortality rates for these children. The incidence is high at 18 percent of these young children in the GLSS5+ to 26 percent among toddlers in treatment households.

Table 5.6 gives the incidence of persons suffering from diarrhea during the same time period. The rate is about 4 percent for children under five, which is about three times higher than the overall incidence among persons in T-C households. Note that the absolute number of persons reporting

diarrhea is small so that the incidence figures could change substantially with a small change in those reporting the problem.

Table 5.3: Distribution of households, by adult composition, locality and presence of children (%)

Adults in Households	From GLSS5+ (Rural Localities)		From HH Survey (All HH's)			
			Control		Treatment	
	With Children	Without Children	With Children	Without Children	With Children	Without Children
At least one adult of each sex	74.7	32.6	86.3	44.7	86.1	66.7
One adult male	52.3	64.4	49.0	59.2	47.0	58.0
At least two adult males	24.8	8.9	39.1	25.0	40.8	25.9
One adult female	61.3	41.2	46.8	40.8	48.2	49.4
At least two adult females	56.2	18.2	51.3	19.7	50.1	33.3

Note: Adult is anyone 18 or older; columns do not add to 100 because categories are not mutually exclusive.

Table 5.4: Population aged 15 years and older, by educational attainment and sex (%)

	GLSS5+		Household Survey					
			Control			Treatment		
	Male	Female	Male	Female	Total	Male	Female	Total
Never Been to School	23.2	43.0	32.4	44.3	38.7	29.5	47.1	38.9
Less Than MSLC/BECE	31.5	30.2	25.5	29.4	27.6	32.3	26.2	29.0
MSLC/BECE/VOC	31.2	21.1	31.2	22.4	26.5	27.6	23.7	25.5
Secondary or Higher	14.1	5.8	10.9	3.8	7.1	10.4	3.0	6.4
Other	--	--	0.1	0.1	0.1	0.2	0.0	.01
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

MSLC is Middle School Leaving Certificate; BECE is Basic Education Certification Examination; VOC is Certificate from a Vocational Institute.

Table 5.5: Incidence of people suffering from illness or injury during the previous two weeks, by age group and sex (%)

Age Group	From GLSS5+ (Total)			From HH Survey					
				Control			Treatment		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	18.0	19.0	18.0	20.5	16.0	18.4	29.1	22.6	25.7
5-11	8.9	8.5	8.7	13.8	13.2	13.5	13.4	16.4	14.8
12-19	7.2	8.9	8.0	9.0	8.8	8.9	10.4	8.3	9.4
20-49	11.0	15.0	13.0	10.4	13.8	12.2	14.9	14.0	14.4
50+	19.0	25.0	23.0	23.4	28.5	26.1	19.7	26.0	23.1
Total	**12.0**	**15.0**	**13.0**	**14.1**	**15.1**	**14.6**	**16.2**	**16.2**	**16.2**

Table 5.6: Incidence of people suffering from diarrhea during the previous two weeks, by age group and sex (%)

Age Group	From HH Survey					
	Control			Treatment		
	Male	Female	Total	Male	Female	Total
0-4	5.0	2.8	4.0	3.9	3.2	3.6
5-11	1.7	1.6	1.6	1.5	1.6	1.6
12-19	0.6	0.3	0.4	1.1	0.3	0.7
20-49	0.9	0.6	0.7	1.0	0.7	0.8
50+	1.7	1.5	1.6	0.4	1.4	1.0
Total	**1.7**	**1.1**	**1.4**	**1.4**	**1.2**	**1.3**

We can explore how severe the reported illnesses and injuries were by looking at the share of the sick and injured who were not able to continue with their regular daily activities. The entries in Table 5.7 indicate that the majority of those reported to be ill were sufficiently sick that they were unable to continue their regular lives. The pattern is remarkably similar across age groups. Again, we note that the absolute number of persons in many of the table cells is small.

The type of treatment sick persons receive is summarized in Table 5.8. There are substantial differences in the incidence of use of Oral Rehydration Therapy and medicines received from a clinic or doctor, with control communities having much higher incidences in both cases. It is unclear what may be behind this pattern. A substantial share of the sick visited a health facility—33 percent for control communities and 52 percent for treatment communities. The great majority of those visited only did so on a single day. As one would expect much smaller shares of the sick were admitted to a health centre or hospital—4 and 8 percent, respectively for control and treatment communities.

Poor hygiene is a common cause of stomach illnesses, and the survey asked a series of questions on this topic, and the results are displayed in Tables 5.9 and 5.10. Overall practices are not very strong. Households have a low incidence of facilities to improve hygiene such as a place to wash one's hands or a soak-away pit (Table 5.9, top panel). In only about one-third of households did anyone attend a hygiene promotion event in the past year. When asked about how the youngest child's stools are disposed, only about 40 percent reported putting them in a toilet facility, 29 percent went into the trash (Table 5.9, last panel). The respondent's own hygienic practices are mixed (Table 5.10). While nearly two-thirds said they always wash their hands with soap after using the toilet, only about 40 percent reported always washing their hands before eating or preparing food.

Table 5.7: Proportion of people suffering from an illness or injury who had to stop their usual activity two weeks preceding the interview, by locality, age group, and sex (%)

	From GLSS5+ (Rural Localities)			From HH Survey (All HH's)					
				Control			Treatment		
Age Group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	65.8	67.6	66.7	66.7	57.1	63.2	70.0	62.5	66.7
5-11	71.7	64.2	68.3	57.1	60.0	58.3	62.5	40.0	53.8
12-19	64.1	70.1	67.3	0.0	50.0	20.0	50.0	0.0	33.3
20-49	64.8	68	66.7	75.0	75.0	75.0	28.6	50.0	36.4
50+	65.7	64.8	65.1	100.0	75.0	87.5	0.0	100.0	75.0
Total	**66.2**	**66.8**	**66.6**	**63.3**	**63.6**	**63.5**	**53.3**	**54.5**	**53.8**

Note: Denominator is those with water-borne illnesses (watery diarrhea, guinea worm, bilharzias).Information was not collected on stopping the usual activities for other illnesses or for injuries.

Table 5.8: Care received by sick persons

Type of Care	Percent of Sick Persons	
	Control	Treatment
Treatments Received		
Boiled Water	3.8	1.9
No Feeding (Apart from Breastfeeding)	0.0	0.0
Fed Easily Digestible Food (1)	0.0	0.0
Breastfeeding	0.0	0.0
Home-Made Oral Rehydration Therapy (ORT) Fluids Using Safe Water (Boiled)	5.8	7.7
Oral Rehydration Therapy (ORT) Sachets Using Safe Water (Boiled)	11.5	1.9
Less Water Than Usual	0.0	0.0
Zinc Supplement	0.0	0.0
Traditional Remedies Such as Cold Tea and Pumpkin Seeds	0.0	0.0
Other Traditional Remedy	11.5	3.8
Medicines from Clinic/Doctor (Other Than Those Above)	65.4	15.4
Extra Water Without Boiling	0.0	0.0
No Treatment	5.8	5.8
Visited Health Facility (a)	**32.7**	**51.9**
Number of Days Spent at Facility		
1-2	82.3	74.0
3-4	11.8	22.2
5 or more	5.9	3.7
Admitted to Health Centre or Hospital (a)	**4.1**	**7.8**
Number of Days Spent at Facility		
1-2	0.0	75.0
3-4	50.0	25.0
5 or more	50.0	0.0

Note: Small N's are the basis for these figures. Small changes in counts result in large percentage changes.

(a): This is percent of sick people with guinea worm, bilharzias, or watery diarrhea who visited a health facility, and the distribution of days spent there.

Table 5.9: Household hygiene

Hygienic Practices of Household	Percent of HH's	
	Control	Treatment
Household Hygiene Facilities and Practices (Y/N)		
HH Has Handwashing Facility, Special Basin	16.0	15.5
HH Has Soak-Pit (Soak-Away Pit)	3.8	3.3
HH Has Rubbish Pit	24.3	23.3
Participation in Hygiene Promotion (Y/N)		
HH Participated in Hygiene Promotion in the Past Year	31.0	32.7
There Have Been Visits to the Community to Follow Up on Hygiene Promotion	17.7	19.0
Means of Disposal of the Youngest Child's Most Recent Stools		
Child Used Toilet/Latrine	15.5	14.8
Put/Rinsed into Toilet or Latrine	24.0	19.8
Put/Rinsed into Drain or Soak-Away Pit	0.7	1.0
Thrown into Garbage	29.4	29.2
Buried	8.5	11.2
Left in the Open	0.7	0.8
Other	21.2	23.2
Total	**100**	**100**

Table 5.10: Respondent Hygiene

Hygienic Practices	Percent of Respondents	
	Control	Treatment
How often do you feel you need to wash your hands with soap after using a toilet?		
Always	64.7	65.8
Sometimes	25.3	24.8
Never	10.0	9.3
Total	100	100
How often do you feel you need to wash your hands with soap before eating?		
Always	41.3	41.7
Sometimes	35.5	34.0
Never	23.2	24.3
Total	100	100
How often do you feel you need to wash your hands with soap before preparing food?		
Always	37.8	39.3
Sometimes	32.8	28.8
Never	29.3	31.8
Total	**100**	**100**

The next block of tables deals with household members' employment over the past 12 months. Table 5.11 provides information on gainful employment using a very broad definition of employment as essentially any activity that generates income, in cash, in-kind, or imputed. The table shows the age distribution of these workers. The patterns are essentially the same for the GLSS5+ and the water survey, with workers concentrated in the 25-44 age bracket.

The great majority of households have multiple workers. Indeed, only about 20 percent have a single worker (Table 5.12). About 40 percent have two workers and 16 percent three workers. Each of two occupations account for the main occupation of one-third of workers: any type of self-employment and working on the family plot, food garden or cattle post (Table 5.13). Another 21 percent are engaged in collecting vegetables or catching animals for sale or for home consumption. The patterns are identical for treatment and control communities.

Table 5.11: Current economic activity rate, by sex, age group and locality (%)

Age Groups	From GLSS5+ (Rural Localities)		From HH Survey (All HH's)			
			Control		Treatment	
	Male	Female	Male	Female	Male	Female
7-14	6.6	4.4	4.2	3.4	4.7	3.3
15-24	16.0	14.9	19.2	17.0	19.4	15.4
25-44	44.0	45.6	44.3	48.5	43.1	50.2
45-64	25.2	27.5	23.1	24.0	23.2	23.8
65 and above	8.1	7.6	9.1	7.1	9.6	7.4
Total	**100**	**100**	**100**	**100**	**100**	**100**

Note: "Economic activity" includes any work for pay, profit, family gain, or production of anything for barter or home use during the past 12 months.

Table 5.12: Number of workers per household and types of occupation (%)

Number of Workers	Percent of HH's	
	Control	Treatment
0	2.0	1.5
1	21.0	20.8
2	41.3	37.3
3	15.8	16.7
4	7.7	8.8
5	4.3	7.0
6	3.8	2.8
7+	4.1	5.1
Total	**100**	**100**

Table 5.13: Main Occupation of Working Household Members (percent of those working)

Type of Main Occupation	Percent of All Working HH Members	
	Control	Treatment
Any Kind of Business, Big or Small, for Himself/Herself (1)	34.5	36.1
Help Unpaid in a Family Business of Any Kind (2)	3.1	3.3
Help on the Family Plot, Food Garden, Cattle Post (3)	33.4	33.0
Collect Vegetables or Catch Animals or Other Food for Sale or as Family Food	21.3	21.3
Do Any Work for a Wage, Salary, or Any Payment in Kind, Even if Only for One Hour	7.7	6.2
Beg for Money or Food in Public	0.1	0.1
Total	**100**	**100**

1. Examples: Selling things, making things for sale, repairing things, guarding cars, donkey cart or other transport business etc.
2. Examples: Help to sell or make things for sale, cleaning up for a business, etc. Don't count normal househork
3. Examples: Ploughing, harvesting, weeding, looking after livestock

Table 5.14 gives information on the types of employers of those workers who work for wages, salaries, or any in-kind payment. Note that the sample size is small, about 100 overall. Hence, focusing on the major categories of public sector, private sector, and other is recommended. The broad pattern in the data from the household survey is for the private sector to be the dominant employer followed by the public sector, and then "other." Interestingly, a substantial share, 25-35 percent, of these workers is employed in the public sector (figures from columns for the household survey, gender disaggregated columns).

Table 5.14: Type of employer for the currently employed population aged 15-64 years, by locality and sex (%)

	GLSS5+ (rural localities)			From HH Survey (all HHs)					
				Control			Treatment		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Public sector	**39.0**	**44.6**	**40.5**	**37.3**	**24.0**	**34.2**	**35.2**	**36.3**	**35.7**
Civil service	21.1	23.6	21.8	12.8	12.0	12.6	14.7	12.1	13.9
Other public service	17.4	20.2	18.1	23.3	12.0	20.7	17.6	21.2	18.8
Parastatals	0.5	0.8	0.6	1.2	0.0	0.9	2.9	3.0	3.0
Private sector	**55.2**	**50.8**	**54.0**	**48.8**	**64.0**	**52.2**	**57.4**	**42.4**	**52.5**
Formal	17.5	10.1	15.6	18.6	20.0	18.9	20.6	12.1	17.8
Informal	37.7	40.7	38.4	30.2	44.0	33.3	36.8	30.3	34.7
Others	**5.8**	**4.6**	**5.5**	**14.0**	**12.0**	**13.5**	**7.4**	**21.2**	**11.9**
NGOs	0.2	0.3	0.2	1.2	0.0	0.9	1.5	3.0	2.0
Cooperatives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
International organisations	0.7	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0
Agric business	3.8	4.0	3.9	10.5	12.0	10.8	5.9	9.1	6.9
Other	1.1	0.3	0.9	2.3	0.0	1.8	0.0	9.1	3.0
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Note: Table calculated using main occupation. We only asked this question if a respondent said their occupation status was "Do any work for a wage, salary, or any payment in kind, even if only for one hour". N's for women are 25-30 and N's for men are 70-85, making total N's about 100.

What do workers earn? The questionnaire asked about both cash and in-kind payments (Table 5.15). Nearly 90 percent of workers received cash payments for their work, with a smaller share getting in-kind payments, about 16 percent for main occupations and 21 for secondary jobs. About 40 percent of workers received GHc or less in cash per month from their primary jobs; the parallel figure for secondary jobs is about 50 percent. Only 16-18 percent had cash payments over GHc 170 per month for both job types.

In-kind payments are often significant. About 60 percent of workers received goods valued at under GHc 30 per month for both primary and secondary positions. But 12-18 percent had in-kind payments in the GHc70-170 range and for around 6-8 percent the value was over GHc 170 among those getting such payments.

Table 5.15: Payments for work in main and secondary occupations

	Percent of All Working HH Members Main Occupation		Percent of All HH Members with a Secondary Occupation	
Main Occupation	**Control**	**Treatment**	**Control**	**Treatment**
Received or Will Receive Money for Work (% Yes)	**84.9**	**86.5**	**89.6**	**87.1**
Monthly Payment Equivalent (GHc)(a)				
0-10	19.5	21.1	28.0	26.7
10.1-30	20.8	23.0	22.2	19.5
30.1-70	21.4	22.0	17.4	19.5
70.1-170	19.5	18.3	16.5	17.4
More than 170	18.8	15.6	15.9	16.9
Receives In-Kind Payment (% Yes)	**16.0**	**16.9**	**20.9**	**21.0**
Monthly Value of In-Kind Payment				
0-10	32.9	48.7	35.4	54.4
10.1-30	27.3	19.4	18.8	8.8
30.1-70	20.1	13.6	18.7	22.8
70.1-170	12.1	11.8	18.8	10.5
More than 170	7.6	6.5	8.3	3.5

a. Percent of those receiving monetary payment for their main occupation. Other groups of distributions in this table follow the same logic.

Data on household income from all sources, in cash and in-kind, are presented in Table 5.16. Note that the survey did not ask about some sources such as interest income, income from rental or property or remittances received from members of their extended families. So the values we report may be lower than from some other sources that employ more comprehensive measures. For total household income, about 40 percent have under GHc 100 per month and the same percentage over GHc 200 per month. On a per capita basis, about 40 percent of households are living on GHc 35 per month or less and another 20 percent on GHc 35-80 per month.

Our final indicator of economic status is type and number of durable goods owned by households in the T-C communities (Table 5.17). The survey inquired about eight commonly owned goods. Radios or cassette players are by far the most frequently owned, held by two-thirds of households. Second place, with about 50 percent ownership, is a bicycle; and televisions with a 25 percent ownership rate are in third position.

The fact that 16 percent of households own none of these durable goods is a strong signal of the depth of poverty for a part of the population in the communities surveyed. The modal number of these durables owned is between 1 and 2. Only 18 percent of households own more than three.

Table 5.16: Monthly incomes per working household member and per household

Monthly Income per Person	Percent of All Working HH Members	
	Control	Treatment
0-15	22.4	24.0
15.1-35	17.4	19.8
35.1-80	19.9	19.4
80.1-200	21.9	20.9
More than 200	18.4	15.9
Monthly Income per Household	**Percent of All Households**	
0-30	14.3	19.2
30.1-100	26.9	23.3
100.1-200	16.0	15.5
200.1-500	23.1	25.0
More than 500	19.7	17.0

Note: Income includes both cash and in-kind payments.

Table 5.17: Durable goods ownership by households

Durable Good	Percent of Households	
	Control	Treatment
Electric Iron	14.8	13.2
Refrigerator	11.8	8.2
Television	24.7	27.7
Video Deck	16.8	19.2
Cassette Player/Radio	66.5	63.2
Sewing Machine	19.0	24.2
Bicycle	47.2	48.3
Number of Durable Goods Owned		
0	16.3	17.3
1	27.3	26.8
2	25.0	21.7
3	13.8	15.5
4	9.2	10.0
5	5.2	5.7
6	2.7	2.2
7	0.5	0.8

We now turn to the type and quality of dwellings occupied by the households in the surveyed communities. A sizable majority of households in the surveyed communities —about 75 percent--live in rooms within a larger structure that is shared with other room owning or renting occupants (Table 5.18). Most of these dwellings are compound houses. The next most frequent arrangement is the occupancy of multiple huts within a compound. The patterns are broadly the same in data from the GLSS5+ and the water survey, with differences likely attributable to the GLSS5+ data being for all areas, not just rural zones.

Eighty percent of households surveyed own the unit in which they live (Table 5.19). About 10 percent are rent paying occupants and 5 percent occupy units rent free.

Table 5.20 shows that there is substantial variance in the number of rooms occupied by a household. Fully 24 percent of households occupy a single room and another 26 percent have two rooms. At the other end of the scale about 17 percent occupy five or more rooms.

There are also sharp differences in the quality of dwellings as indicated by the durability of the materials from which they are constructed (Table 21). Households are evenly balanced between those having units with walls made of mud or mud bricks and those with walls constructed from cement or sandcrete blocks.

In contrast, the incidence of durable materials is greater for floors and roofs. Floors are of particular interest because analyses have shown that cement floors have very large positive effects on toddler health and children's school performance (Cataneo et al., 2009). In the communities surveyed 78 percent of the dwellings occupied have cement or concrete floors. Because of the Ghana's heavy rains, a strong roof is also an important contributor to good health. Here the incidence of strong materials, corrugated iron sheets, at 60 percent of units is less than for durable flooring. Clearly the analysis of health improvements associated with improved water services will have to control for the quality of respondents' walls, floors, and roofs.

Table 5.18: Households, by type of dwelling (%)

Dwelling Type	From GLSS5+ (All MiDA Zones)	From HH Survey (All HH's)	
		Control	Treatment
Separate House (Bungalow)	2.7	1.2	1.7
Semi-Detached House	2.9	2.3	1.7
Flat/Apartment	3.2	0.3	0.7
Rooms (Compound House)	41.1	53.2	49.8
Rooms (Other Type)	26.0	24.8	28.8
Several Huts/Buildings (Same Compound)	18.1	11.5	11.0
Several Huts/Buildings (Different Compounds)	2.9	6.7	6.0
Tents/Improvised Home	0.1	0.0	0.3
Other	3.2	0.0	0.0
All	**100**	**100**	**100**

Table 5.19: Households, by present occupancy status (%)

Occupancy Status	From GLSS5+ (All MiDA Zones)	From HH Survey (All HH's)	
		Control	Treatment
Owning	55.4	81.2	85.8
Renting	12.3	12.0	8.5
Rent Free	32.1	5.5	3.8
Perching*	0.2	1.3	1.7
Other	--	0.0	0.2
Total	**100**	**100**	**100**

*Unofficial co-occupancy

Table 5.20: Households by number of rooms occupied (%)

Number of Rooms	From GLSS5+ (All MiDA Zones)	From HH Survey (All HH's)	
		Control	Treatment
1	39.2	24.8	23.2
2	28.0	27.5	25.3
3	14.8	18.2	19.7
4	7.3	13.5	14.2
5+	10.8	16.0	17.6
Total	**100**	**100**	**100**

Table 5.21: Main construction materials used by households (%)

Materials	From GLSS5+ (All MiDA Zones)	From HH Survey (All HH's)	
		Control	Treatment
Outer Wall Materials			
Cement/Sandcrete Blocks	48.2	38.8	34.2
Mud/Mud Bricks	47.0	56.8	59.2
Other	4.8	4.4	6.6
Total	**100**	**100**	**100**
Floor Materials			
Cement/Concrete	73.4	77.8	78.7
Earth/Mud/Mud Bricks	22.3	21.0	19.8
Other	4.3	1.2	1.5
Total	**100**	**100**	**100**
Roof Materials			
Corrugated Iron Sheet	56.1	60.0	61.3
Palm Leave/Raffia/Thatch	30.7	33.0	32.2
Other	13.2	7.0	6.5
Total	**100**	**100**	**100**

As a bridge to detailed tables on water sources, access to it and its cost, the next block of tables presents information on households' main sources of water and the type of toilet facilities they use.

Tables 5.22 and 5.23 give the percent distribution of respondents' water sources for drinking and cooking and for general usage, respectively. Two points stand out immediately. First, the sources for the two types of use are essentially identical. Second, the pattern for water survey respondents differ significantly from that in the GLSS5+ data, which we attribute to the GLSS data being for both urban and rural areas.

Sixty percent of respondents in both treatment and control communities get their water from wells broadly defined. But there is considerable diversity between the two community types for other sources, with households in control communities more often using pipe-borne water (15 to 6 percent for treatment communities) and less often using rain and surface water which is often of low quality (21 to 33 percent for treatment communities). These patterns confirm that MiDA is targeting communities with comparatively dire water problems with its investments.

Table 5.22: Households by main source of water for drinking and cooking (%)

Main Source of Drinking Water	From GLSS5+ (All MiDA Zones)	From HH Survey (All HH's)	
		Control	Treatment
Pipe-Borne	**34.9**	**15.1**	**5.7**
Indoor Plumbing	1.3	0.5	0.0
Inside Standpipe	6.4	--	--
Pipe in Neighboring Household	8.2	1.5	1.1
Private Outside Standpipe/Tap	5.5	0.6	0.3
Public Standpipe	13.6	12.4	4.3
Well	**34.5**	**59.3**	**57.6**
Borehole	24.4	47.5	45.1
Protected Well	5.1	2.2	5.7
Unprotected Well	4.9	9.7	6.8
Natural Sources	**26.6**	**21.2**	**33.0**
Rain Water/Spring	2.3	2.6	2.5
Surface Water (Dugout/Pond/Lake/Dam/River/Stream)	24.3	18.6	30.4
Other	**4.1**	**4.5**	**3.7**
Water Truck/Tanker Service/Water Vendor	2.1	1.7	1.0
Cart with Small Tank/Drum	--	0.2	0.8
Sachet/Bottled Water	1.6	1.5	1.6
Other	0.3	1.1	0.3
Total	**100**	**100**	**100**

Note: Main source identified by the source used for "drinking and cooking" from which the household collects the most water in one week. If there are two main sources from which the same volume is gathered, one was chosen at random.

Table 5.23: Households by main source of water for general uses (%)

Main Source of Water for General Use	From GLSS5+ (All MiDA Zones)	From HH Survey (All HH's)	
		Control	Treatment
Pipe-Borne	**29.3**	**14.6**	**5.5**
Indoor Plumbing	1.0	0.4	0.0
Inside Standpipe	6.5	--	--
Pipe in Neighboring Household	6.5	1.5	1.1
Private Outside Standpipe/Tap	4.8	0.6	0.3
Public Standpipe	10.5	12.1	4.1
Well	**36.6**	**58.9**	**55.0**
Borehole	22.0	45.9	42.3
Protected Well	7.5	2.7	6.1
Unprotected Well	7.2	10.3	6.7
Natural Sources	**31.7**	**22.2**	**35.8**
Rain Water/Spring	2.4	2.5	2.6
Surface Water (Dugout/Pond/Lake/Dam/River/Stream)	29.3	19.7	33.2
Other	**2.4**	**4.3**	**3.8**
Water Truck/Tanker Service/Water Vendor	2.1	1.8	0.9
Cart with Small Tank/Drum	--	0.1	0.9
Sachet/Bottled Water	--	1.3	1.5
Other	0.3	1.0	0.5
Total	**100**	**100**	**100**

Note: Main source of water defined as being the source from which the most water is collected by household members in one week during the dry season. If there are two main sources from which the same volume is gathered, one was chosen at random.

Regarding the toilet facilities used by households in the water survey, only 38 and 33 percent, respectively, of those in control and treatment communities, used improved facilities (Table 5.24). About 30 percent in both community types use no type of facility whatsoever.

Lastly, Table 5.25 provides information on the development of toilet facilities, showing why type of facilities have been sponsored by the government or NGOs, those constructed by private owner without assistance, those constructed by the private owner getting a government subsidy and a catch-all "other" category. In both treatment and control communities the government and NGOs have favored three solutions: the flush pit latrine, the ventilated improved pit latrine, and the pit latrine without a slab (open pit). Individual owners have invested in a wider range of solutions.

Table 5.24: Households by type of toilet used (%)

Type of Toilet	From HH Survey (All HH's)	
	Control	Treatment
None	**29.8**	**33.0**
Improved	**38.3**	**32.9**
Flush/pour to…		
…piped sewer system	1.2	0.3
…septic tank	1.2	0.5
…pit latrine	1.2	2.3
…elsewhere	0.7	0.8
…don't know	0.0	0.0
Ventilated improved pit latrine	6.5	6.0
Pit latrine with slab	28.3	23.7
Unimproved	**10.4**	**8.3**
Pit latrine without slab (open pit)	9.7	7.3
Pan/bucket	0.0	0.2
Public facility	**21.5**	**25.8**
Total	**100**	**100**

Table 5.25: Means of building and paying for toilets, Households in MiDA survey

Toilet Type	Control					Treatment				
	Owner Built, Family Savings and Support	Owner Built, Subsidy from Govt or NGO	Govt/NGO Built	Other	Total	Owner Built, Family Savings and Support	Owner Built, Subsidy from Govt or NGO	Govt/ NGO Built	Other	Total
Flush/pour to…										
…piped sewer system	85.7	0.0	14.3	0.0	100	50.0	0.0	50.0	0.0	100
…septic tank	100.0	0.0	0.0	0.0	100	100.0	0.0	0.0	0.0	100
…pit latrine	50.0	33.3	16.7	0.0	100	71.4	7.1	21.4	0.0	100
…elsewhere	0.0	0.0	0.0	100.0	100	0.0	0.0	0.0	100.0	100
Ventilated improved pit latrine	53.8	17.9	17.9	10.3	100	63.9	19.4	8.3	8.3	100
Pit latrine with slab	55.3	9.4	8.8	26.5	100	53.5	5.6	5.6	35.2	100
Pit latrine without slab (open pit)	44.8	10.3	24.1	20.7	100	29.5	4.5	31.8	34.1	100
Pan/bucket	--	--	--	--	100	100.0	0.0	0.0	0.0	100
Public facility-VIP/KVIP	11.6	0.8	73.6	14.0	100	5.8	1.9	74.2	18.1	100

The presentation now turns to the details of water acquisition, consumption and cost. Five introductory points are in order before turning to the tables. First, we generally present data for the control and treatment populations combined. This is done because the patterns for the two groups are very similar. With complex tables comparing the values of the two groups is very difficult. The simplification facilitates seeing the main patterns.

Second, in the detailed tables where results are presented for each water source type, we omit those water sources used by fewer than 20 households. All sources are, however, included in less complex tables where results are summed over water source categories.

Third, there are no comparisons with GLSS5+ data in this section because of the very large differences in the questions on water between our survey and the GLSS5+. For time spent fetching water, for example, GLSS5+ asks a question of each household member about how much time they spend collecting water, asking separately for times when the respondent is caring for children and not caring for them (Section 4, Part g, time use). Our survey, on the other hand, asks separately about the time spent fetching water, source used by the household. Moreover, our questions asked about time spent in each stage of the collection process: walking to/from the source, waiting in line to obtain water, and time actually obtaining the water. The greater detail in our survey is what one expects in one that is focused on water use. But it does mean that the data are essentially non comparable.

Fourth, there were a small number of extremely large values recorded for key data items, e.g., time spent walking to a water source, container size. We developed rules for identifying extreme values and replacing them with reasonable large but not extreme values. These are described in Annex C.

Fifth, the questionnaire concentrated on water sources used in the dry season because the water investment project will yield its greatest benefits during this season. The information presented reflects this orientation.

The first set of tables reviewed provides information on access and related factors by type of water source. They provide data on individuals and households collecting and using specific water sources. The second set then combines the information on the water sources used by each household to generate comprehensive information at the household level. As we will see, the typical household uses 2-3 different sources; so the second set of tables should be used to understand patterns at the household level.

Table 5.26 provides information on water acquired during the dry season, or both seasons if it is acquired in both, for drinking and cooking purposes. Data are given for each water source separately. Several sources are used significantly in the dry season only. About 29 percent of households use

water from someone else's yard only during the dry season and 48 percent only use water from tanker trucks then. Three more sources are only used by 18 percent of households in dry weather.

Respondents were asked if the water from each source is safe to drink in the dry season. Three sources are rated by 38-45 percent of households as unsafe: unprotected dug wells, tanker trucks, and surface water, e.g., a lake. On the other hand, tube wells and public standpipes were rated safe by nearly everyone.

There is a number of sources where the share of household treating the water for drinking is significantly smaller than the share saying it is unsafe to drink: while only 40 percent of respondents said that unprotected dug wells yield unsafe water, 56 percent report never treating it; the parallel figures for tanker truck water are 38 and 54 percent. In terms of treatment, boiling and adding bleach/chlorine are common methods across sources. Another common treatment is just letting the water stand and settle, which may not be very effective treatment for drinking water.

Similar information is provided in Table 5.27 for water used for general purposes. The most striking feature of this table is how similar the entries are to those for drinking and cooking water in Table 5.26.

Households devote substantial time daily in the dry season to acquiring water for essential purposes, as documented in Table 5.28. Note that the upper part of the table reports information for individuals collecting water and the lower part aggregates data for individual collectors into household values.

The upper panel shows the time in minutes devoted by one person to acquiring water during one visit to a source for each of the 11 sources included in the table. Information was obtained on time spent on each of the three distinct steps in the acquisition process: going to/from the water sources, time waiting in line to collect the water, and the time to collect the water, e.g., to draw water from a well.

Water sources fall into three groups based on the total time required.

- There are six sources with high time requirements per trip, i.e., times approaching one hour. These are the first seven sources in the table with the exception of a public standpipe which has lower times. Average time to go to/from a source is 16-30 minutes, while waiting time ranges from 21 to 37 minutes. Actually obtaining the water averages 4 to 6.5 minutes.

- The second group has three sources with total average times in the 5 to 18 minute range: collected rainwater, bottled water and tanker trucks. Times going to/from sources are very short except for tanker trucks (11 minutes).

- The final group of three sources is in an intermediate position for total average acquisition time: public standpipes, carts with a small tank or drum, and surface water. Total average time ranges from 25 to 39 minutes.

The typical person collecting water makes 12-15 trips per week for the first seven sources in the table plus surface water. But the trip numbers for the three low total time sources and the cart-with-drum are in the 6-10 range, perhaps indicating these are supplemental rather than primary sources.

One can contrast the number of persons collecting water from a source with the number of households using each water source. Typically about 1.5 persons per household are collecting water from a source. For example 245 persons from 145 households collect water from a public standpipe, or about 1.7 persons per household.

When one aggregates the trips taken by all household members, one gets the total minutes per household per day that households spend getting water from a source. The highest value is 270 minutes per day for water from an unprotected spring. The mean values for four sources are around 200 minutes per day: getting water from someone else's plot, a tube well/borehole, an unprotected well, and a protected spring. The values for collected rainwater and bottled are dramatically lower. It is important to look at the distribution of household over the various time categories in the table because they show that while the mean values are often high, large minorities and often the majority of households using a source spent less than 60 minutes a day at the task.

We prepared two summary tables on time spent fetching water to give an overview of the mass of data in Table 5.28 and to show that households in control and treatment communities have broadly similar patterns. These data are for all sources used by each household, i.e., we take account of the fact that households use a variety of water sources and tailor sources to their specific water needs. Table 5.29.1 gives information at the household level and Table 5.29.2 converts the same information to a per capita basis for the number of persons in each household.

Because essentially all households use multiple water sources (data presented below), one anticipates that total times for all sources used will generally be larger than those for individual sources. And this is the case. All household members together are spending around four hours a day (240 minutes) collecting water. There is very little difference between the time collecting water for drinking and cooking purposes and all purposes, indicating that most effort is going to meet essential water requirements. But the distribution of time spent among households is also very important and informative. About 20 percent spend less than 30 minutes a day fetching water, and about half spend under two hours a day. Rather shockingly, about 25 percent have household members devoting over six hours a day to acquiring water.

How much water are households gathering per day from different water sources and what do they pay for it? The information in Table 5.30 gives some answers to these questions. The top panel displays the size of the containers water collectors are using. Many containers are surprisingly large (and heavy when filled)—over 40 liters.[14] While about half of those collecting water from most sources in the table are using containers of 40 liters or less, the balance are using larger ones. Presumably those using such big containers are using some sort of wagon to carry them. Those taking water from highly convenient sources of bottled water and a cart with a water drum are using smaller containers, mostly those under 20 liters. The figures for rain collection are somewhat ambiguous; we believe that in many cases the respondent reported that size of the rainwater collection container, not the size of the container s/he uses in drawing water from it.

The middle table panel provides information on total household water consumption per day. (Per capita values per household from all sources are presented later.) Most mean values are in the 130 – 200 liters per day range. Three sources that may well be tapped for specific purposes or under special conditions have much lower mean values in the 38-70 liters per day range.

The majority of households, often a very large majority, are not paying for water from unimproved sources. For example, nearly 90 percent do not pay for water from unprotected dug wells, unprotected springs, and rainwater collection. In contrast, nearly everyone pays for piped water, water from public standpipes, bottled water, carts with drums, and tanker trucks. The great majority of those paying make payments based on the container size, although around 20 percent pay on a monthly basis. Even for those paying the cost is often not terribly large. Water from a public standpipe has an average cost of GHc 49 per month, water from boreholes averages GHc 63 per month, and GHc 44 for protected dug wells. Those consuming bottled water spend much more: GHc 93 per month on average. The highest monthly expenditures are for households using water from tanker trucks—GHc 188 per month.

Table 5.31 summarizes the information in Table 5.30 by aggregating water sources into two groups: improved and unimproved sources (definitions at the beginning of this chapter). Households are acquiring a larger quantity of water daily from improved that unimproved sources: 216 liters vs. 173. As expected a higher share of households pay for water from improved sources than for unimproved: 50 percent vs. 21 percent. Those paying for water from improved sources pay by the container about twice as often as they pay by the month. Interestingly, those who do pay for water from the different sources, those purchasing from unimproved sources pay more per month on average. Those purchase water from improved sources pay GHc 30 and those from unimproved sources GHc 53.

[14] Interviewers obtained information on container size by reading any information on the container first. If this was not available, the container was measured.

Table 5.32 present the same information as in Table 5.31 but only for water sources used primarily for drinking and cooking. The patterns are very similar.

So what do households pay in a month for water from all the sources they access? Relevant data are in Tables 5.33-5.35. Total spending is computed by weighting the price of water from each source used by a household by the share of total water consumption accounted for by each source. Where a water source is free, a price of zero is used when the share of the source in total household water use in multiplied by the price.

Table 5.33 shows that households in control communities pay GHc 27 and those in treatment communities GHc 48 on average. Because fewer households in treatment communities pay for water from any source, the result indicates that those who do pay are paying higher prices than those in control communities. From the information in Table 5.31, we infer that households in treatment communities may have been more often purchasing water from unimproved sources.

In Table 5.34-5.35 monthly costs are shown for all households, not just those paying for water from at least one source. Table 5.34 gives the information for on a per households basis and Table 5.35 on a per capita, rather than per household, basis.

We now return to a further exploration of the volume of water collected (and consumed). Table 5.36 displays information on the mean water collected and its percentage distribution at the household level. Separate figures are given for households in control and treatment communities and for water collected for all purposes and primarily for drinking and cooking. The figures for controls and treatments are very similar. The data on volumes for all purposes and only drinking and cooking are very close, indicating that vast majority of water being collected is for drinking and cooking purposes.

Table 5.37 presents the same information on a per capita basis. These data are particularly important because they can be compared with international standards for minimum water requirements. For drinking and cooking the international minimum standard is 15 liters per person per day (l/p/d) with a range of 12 to 55 l/d/p (Gleick, 1996). The data in the table show that for households in both control and treatment communities about 22 percent gather 15 l/d/p or less per day for drinking and cooking purposes. But about the same share of households have about the same volume of water collected for all purposes. This is clearly a significant problem.

Mean per capita amounts of water collected are higher—running about 50 l/d/p, with values slightly less in control villages. Hence, the mean is only at the upper end of the minimum daily water requirements.

The discussion of water collection and cost has often referred to multiple water sources used by a household. This concluding part of the discussion of the household data addresses this point. Table 5.38 shows the distribution of the number of sources used by households in control and treatment communities and all households combined. Once again the patterns for the two community types are essentially identical. The typical household accesses 2-3 water sources on a regular basis. About 15 percent use 4 or more sources.

The matrix for all 15 water sources inquired about in the household survey shown in Table 5.39 documents that many combinations of sources are accessed. The table shows that, for example, 6 households reported using piped water into yard/plot (cell 2,2). Of those 6, one also reported using a tube-well/borehole, 4 also reported using rainwater collection, and one also reported using bottled water. The 31 households who reported using water piped into someone else's yard also used water from 11 other sources! Clearly, households have devised complex strategies based on their location relative to various sources, the availability of household members to fetch water and the time and monetary costs of using each water source.

Table 5.26: Water use and treatment, for water sources used for cooking and drinking during the dry season or both seasons

Question	Water Sources									
	Piped into Someone Else's Yard/Plot	Public Tap/ Standpipe	Tube-Well/ Borehole	Protected Dug Well	Unprotected Dug Well	Unprotected Spring	Rainwater Collection	Bottled Water	Tanker-Truck	Surface Water (River, Lake, Pond)
Number of Households Using This Source for Cooking/drinking	28	143	693	83	142	25	44	170	29	449
Seasons Water is Used from This Source (b)										
Dry Season	28.6	18.2	12.4	18.1	18.3	16.0	4.5	8.8	48.3	18.3
Both Seasons	71.4	81.8	87.6	81.9	81.7	84.0	95.5	91.2	51.7	81.7
During the Dry Season, Water is Safe to Drink										
Yes	89.3	94.4	93.4	85.5	59.9	84.0	81.8	100.0	62.1	54.6
No	10.7	5.6	6.6	14.5	40.1	16.0	18.2	0.0	37.9	45.4
Frequency of Water Treatment During the Dry Season (for Water that is Unsafe to Drink)										
Always	33.3	0.0	14.6	14.3	24.6	25.0	25.0	--	36.4	35.7
Most of the Time	0.0	20.0	8.3	0.0	5.3	25.0	0.0	--	0.0	10.6
Sometimes	33.3	10.0	4.2	0.0	14.0	25.0	12.5	--	9.1	11.1
Never	33.3	40.0	72.9	85.7	56.1	25.0	62.5	--	54.5	42.6

Question	Water Sources									
	Piped into Someone Else's Yard/Plot	Public Tap/ Standpipe	Tube-Well/ Borehole	Protected Dug Well	Unprotected Dug Well	Unprotected Spring	Rainwater Collection	Bottled Water	Tanker-Truck	Surface Water (River, Lake, Pond)
Treatments Used During Dry Season (for Water That is Treated)										
Boiling	0.0	20.0	27.3	0.0	20.8	33.3	66.7	--	25.0	28.7
Add Bleach/Chlorine	0.0	0.0	18.2	0.0	29.2	33.3	0.0	--	0.0	8.7
Strain Through Cloth	0.0	20.0	18.2	0.0	16.7	33.3	0.0	--	25.0	47.0
Use Water Filter (Ceramic, Sand, Composite, Etc)	0.0	20.0	9.1	0.0	8.3	0.0	0.0	--	50.0	10.4
Let It Stand and Settle	100.0	20.0	18.2	100.0	16.7	0.0	33.3	--	0.0	3.5
Solar Disinfection	0.0	20.0	0.0	0.0	8.3	0.0	0.0	--	0.0	0.9
Other	0.0	0.0	9.0	0.0	0.0	0.1	0.0	--	0.0	0.8

--Means there were under 20 observations in the cell. Only water sources with at least 20 users are included in the table.

Table 5.27: Water use and treatment, for water sources used for general purposes during the dry season or both seasons

Question	Water Sources									
	Piped into Someone Else's Yard/Plot	Public Tap/ Standpipe	Tube-Well/ Borehole	Protected Dug Well	Unprotected Dug Well	Unprotected Spring	Rainwater Collection	Bottled Water	Tanker-Truck	Surface Water (River, Dam, Lake, Pond)
Number of Households Using This Source	30	145	713	101	165	27	48	172	32	547
Seasons Water is Used from This Source										
Dry Season	26.7	17.9	12.8	19.8	18.2	14.8	4.2	8.7	46.9	23.6
Both Seasons	73.3	82.1	87.2	80.2	81.8	85.2	95.8	91.3	53.1	76.4
During the Dry Season, Water is Safe to Drink										
Yes	83.3	94.5	93.0	77.2	58.2	85.2	77.1	100	62.5	49.5
No	16.7	5.5	7.0	22.8	41.8	14.8	22.9	0	37.5	50.5
Frequency of Water Treatment During the Dry Season (for Water that is Unsafe to Drink)										
Always	20.0	30.0	15.4	12.0	20.3	25.0	18.2	--	33.3	29.6
Most of the Time	0	20.0	9.6	0	4.3	25.0	0	--	0	10
Sometimes	40.0	10.0	3.8	4	11.6	25.0	27.3	--	8.3	12.5
Never	40.0	40.0	71.2	84.0	63.8	25.0	54.5	--	58.4	47.9

Question	Water Sources									
	Piped into Someone Else's Yard/Plot	Public Tap/ Standpipe	Tube-Well/ Borehole	Protected Dug Well	Unprotected Dug Well	Unprotected Spring	Rainwater Collection	Bottled Water	Tanker-Truck	Surface Water (River, Dam, Lake, Pond)
Treatments Used During Dry Season (for Water That is Treated)										
Boiling	33.3	20.0	30.8	66.7	20.8	33.3	60.0	--	25.0	25.4
Add Bleach/Chlorine	0	0	15.4	0	29.2	33.3	0	--	0	11.4
Strain Through Cloth	0	20.0	15.4	0	16.7	33.3	0	--	25.0	49.3
Use Water Filter (Ceramic, Sand, Composite, Etc)	0	20.0	7.7	0	8.3	0	0	--	50.0	9.2
Let It Stand and Settle	66.7	20.0	23.1	33.3	16.7	0	40.0	--	0	3.5
Solar Disinfection	0	20.0	0	0	0	0	0	--	0	0.7
Other	0	0	7.7	0	8.3	0	0	--	0	0.7

--Means there were under 20 observations in the cell. Only water sources with at least 20 users are included in the table.

Table 5.28: Time to collect water during the dry season

	Water Sources										
Question	Piped into Someone Else's Yard/Plot	Public Tap/Standpipe	Tube-Well/Borehole	Protected Dug Well	Unprotect-ed Dug Well	Unprotected Spring	Rainwater Collection	Bottled Water	Cart with Small Tank/ Drum	Tanker-Truck	Surface Water (River, Dam, Lake, Pond)
Number of People Collecting Water from This Source	43	245	1170	156	257	45	70	208	25	43	882
Mean Collection Time per Water Collector per Trip (1)											
Travelling to and From the Water Source	21.7	11.8	18.5	16.2	23.4	30.7	1.8	8.1	8.1	11.1	29.1
Waiting in Line at the Water Source	36.6	18.1	35.7	21.0	36.0	29.3	0.7	0.8	12.3	3.1	6.5
Filling Containers with Water	3.2	2.6	5.7	6.4	6.6	5.7	2.4	0.3	4.3	4.3	3.1
Total Time Spent	61.5	32.5	59.8	43.5	66.1	65.7	4.9	9.3	24.7	18.5	38.6
Mean Number of Trips per Collector per Week (3)	13.8	11.8	14.4	10.7	15.7	12.9	9.6	7.4	5.9	7.4	11.9
Number of Households Collecting Water From This Source	30	145	712	100	165	27	48	169	15	32	545
Average Time Spent for Water Acquisition per HH per Day (2)	184.5	94.9	212.8	117.8	242.2	270.5	7.7	8.1	22.0	27.7	104.8
0-10	20.0	13.8	5.6	19.0	13.3	11.1	75.0	82.2	66.7	40.6	13.2
10.1-60	13.3	53.1	30.6	32.0	29.7	25.9	22.9	16.0	26.7	50.0	44.0
60.1-200	40.0	18.6	32.2	31.0	26.7	25.9	2.1	1.2	6.7	6.3	29.4
200 and up	26.7	14.5	31.6	18.0	30.3	37.0	0.0	0.6	0.0	3.1	13.4
(1) For each household, we calculate the average time a household member spends collecting water from a given source. The following four rows report the mean of this calculation.											
(2) This is computed using total time, i.e. the sum of the time all HH members spend collecting water in a day from a given source. Means are generally skewed upwards by unlikely responses.											
(3) Note: This is the mean number of total trips by household members, NOT the mean number of trips PER household member											

Table 5.29.1: Total household time spent fetching water per day (minutes)

Cost	All Sources		Sources Used for Drinking/Cooking	
	Control	Treatment	Control	Treatment
Mean	**262.6**	**232.9**	**254.8**	**214.7**
% distribution				
0 to 30	22.3	19.6	23.5	20.9
30.1 to 120	31.0	32.6	31.0	32.5
120.1 to 300	22.9	22.0	23.0	21.9
300.1 and up	23.8	25.8	22.5	24.7
Total	**100**	**100**	**100**	**100**

Table 5.29.2: Total household time spent fetching water per day per capita (minutes (1))

Cost	All Sources		Sources Used for Drinking/Cooking	
	Control	Treatment	Control	Treatment
Mean	**48.3**	**42.0**	**46.7**	**39.7**
% distribution				
0 to 10	34.2	32.4	35.9	35.1
10.1 to 25	24.3	24.5	24.4	22.8
25.1 to 60	18.5	20.8	18.1	20.9
More than 60	22.9	22.3	21.6	21.2
Total	**100**	**100**	**100**	**100**

1. The per capita time was calculated for each household, and this variable is summarized in this table

Table 5.30: Water volume and monthly cost, all water sources

	Water Sources										
Question	Piped into Someone Else's Yard/Plot	Public Tap/Stand pipe	Tube-Well/Bore-hole	Protected Dug Well	Unprotect-ed Dug Well	Unprotect ed Spring	Rainwater Collection	Bottled Water	Cart with Small Tank/ Drum	Tanker-Truck	Surface Water (River, Dam, Lake, Pond)
Volume of Water Container Used (Liters) (a)											
0-20	15.9	20.3	18	14.8	15.2	20.0	6.5	94.5	66.7	25.0	15.2
20.1-40	38.6	31.7	31.2	32.9	33.5	44.4	24.7	1.8	25.9	43.2	33.5
40.1-60	15.9	19.9	20.5	23.2	23.6	6.7	16.9	0.9	0.0	13.6	17.8
60.1-80	22.7	14.6	17.5	16.8	16.0	20.0	28.6	0.5	0.0	4.5	19.6
More than 80	6.8	13.4	12.8	12.3	11.8	8.9	23.4	2.3	7.4	13.6	13.9
N (Number of water collectors)	44	246	1172	155	263	45	77	218	27	44	889
Total Water per Day per HH (Liters)											
Mean	159.3	129.2	183.5	150.4	196.9	139.2	309.4	38.3	46.0	70.2	161.6
0-20 (a)	30.0	15.4	11.6	10.0	6.7	11.1	25.5	66.5	53.3	24.1	12.3
20.1-50	10.0	23.8	13.7	19.0	17.7	0.0	17.0	15.0	20.0	20.7	17.2
50.1-100	16.7	23.1	23.7	23.0	20.1	40.7	14.9	10.8	6.7	37.9	24.8
100.1-200	16.7	15.4	22.3	22.0	22.6	29.6	12.8	4.2	20.0	13.8	24.1
More than 200	26.7	22.4	28.7	26.0	32.9	18.5	29.8	3.6	0.0	3.4	21.6
N (Number of HH's)	30	143	708	100	164	27	47	167	15	29	536
Type of Payment (percent of water collectors)(a)											
Do Not Pay	2.3	4.5	55.4	87.7	89.4	100.0	87.3	0.5	0.0	13.2	98.5
Payment by Month	20.9	17.1	16.6	12.3	0.8	0.0	0.0	5.9	100.0	2.6	0.6
Payment by Container	76.8	78.4	28.0	0.0	9.8	0.0	12.7	93.6	0.0	84.2	0.9
Average Monthly Cost Among Paying House	16.5	49.0	62.9	44.2	58.2	--	3.9	93.2	97.5	188.2	27.0
N (Number of households paying for this so	29	136	307	14	17	0	4	161	13	23	10

Notes:

(a) Percent distributions

First panel shows distribution of volumes used by all water-collecting HH members; so if two members collect water from a source, that household generates two entries in the table. Cost per month averages are calculated using only non-zero values.

Table 5.31: Summary of Household Water Consumption and Incidence of Water Payments

Question	Improved Sources	Unimproved Sources
Total Water per Day per HH (Liters)		
Mean	215.9	173.4
Percent distribution		
0-20	12.3	12.8
20.1-50	13.1	16.3
50.1-100	22.1	23.2
100.1-200	21.9	23.3
More than 200	30.7	24.5
N (Number of HH's)	887	760
Type of Payment (percent of water collectors, percent distribution)		
Do Not Pay	50.4	79.2
Payment by Month	15.2	1.4
Payment by Container	34.4	19.4
Average Monthly Cost Among Paying Households	**29.7**	**52.9**
N (Number of households paying for this source)	467	196

Note: First panel shows distribution of volumes used by all water-collecting HH members; so if two members collect water from a source, that household generates two entries in the table.

Table 5.32: Water volume and monthly cost, water sources used mainly for cooking/drinking

Question	Water Sources	
	Improved Sources	Unimproved Sources
Total Water per Day per HH (Liters)		
Mean	214.7	165.7
Percent distribution		
0-20	12.0	14.8
20.1-50	13.9	15.9
50.1-100	21.9	23.1
100.1-200	21.7	22.9
More than 200	30.5	23.3
N (Number of HH's)	866	681
Type of Payment (percent of water collectors; percent distribution)		
Do Not Pay	49.5	77.0
Payment by Month	15.6	1.6
Payment by Container	34.9	21.4
Average Monthly Cost Among Paying Households	30.3	53.3
N (Number of households paying for this source)	458	193

Note: First panel shows distribution of volumes used by all water-collecting HH members; so if two members collect water from a source, that household generates two entries in the table. Cost per month averages are calculated using only non-zero values.

Table 5.33: Incidence and Amount of Water Payments by Households in Control and Treatment Communities

Question	All Water Sources	
	Control	Treatment
Type of Payment (percent distribution)		
Do Not Pay	58.5	68.3
Payment by Month	10.4	7.1
Payment by Container	31.1	24.5
Average Monthly Cost per Source Among Paying Households	27.3	48.5
N (Number of households paying for water)	323	251

Table 5.34: Total household water cost, All Households, per month (GHc)

Cost	Percent of HH's			
	All Sources		Sources Used for Drinking/Cooking	
	Control	Treatment	Control	Treatment
Mean	**19.0**	**25.9**	**18.4**	**23.3**
% distribution				
0	45.3	57.8	45.3	58.0
.1 to 3	21.7	11.8	21.5	12.2
3.1 to 10	10.8	11.1	11.0	11.3
10.1 to 30	10.7	7.1	10.7	6.8
More than 30	11.5	12.3	11.5	11.7

Table 5.35: Total household water cost, All Households, per capita per month (GHc)

Cost	Per capita cost (1)			
	All Sources		Sources Used for Drinking/Cooking	
	Control	Treatment	Control	Treatment
Mean	**4.3**	**6.2**	**4.2**	**5.9**
% distribution				
0	45.3	57.8	45.3	58.0
0.1 to 1	25.4	17.5	25.3	18.0
1.1 to 3	10.2	8.7	10.1	8.6
3.1 to 8	8.1	4.4	8.5	4.4
More than 8	11.0	11.6	10.8	11.0

1. Per capita cost was calculated for each household with water payments, and this variable is summarized in this table. Households included in the “all sources” and “drinking/cooking” groups differ.

Table 5.36: Total household water volume collected per day

Cost	Percent of HH's			
	All Sources		Sources Used for Drinking/Cooking	
	Control	Treatment	Control	Treatment
Mean	**272.6**	**273.7**	**260.1**	**252.4**
% distribution				
0 to 50	19.7	19.0	21.1	20.4
50.1 to 150	29.5	36.3	29.8	36.7
150.1 to 350	27.2	25.5	27.3	25.9
More than 350	23.6	19.2	21.8	17.0

Table 5.37: Total household water collected per day per capita

Cost	Per capita volume (1)			
	All Sources		Sources Used for Drinking/Cooking	
	Control	Treatment	Control	Treatment
Mean	**53.5**	**49.8**	**51.4**	**46.5**
% distribution				
0 to 15	29.4	32.1	31.3	33.8
15.1 to 25	15.2	17.8	15.2	18.7
25.1 to 75	35.8	34.1	34.8	33.6
More than 75	19.7	16.0	18.7	13.9

1. The per capita volume was calculated for each household, and this variable is summarized in the table

Table 5.38: Number of sources used per household

Number of Sources	Number of Households		
	Control	Treatment	All
0	1	0	1
1	62	43	105
2	186	200	386
3	233	236	469
4	96	96	192
5-7	22	25	47
Total	**600**	**600**	**1200**

Table 5.39: Number of Households Using Various Combinations of Water Sources

	Water Source

	Water Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Piped Water into Dwelling	4														
2	Piped Water into Yard/Plot	0	6													
3	Piped into Someone Else's Yard/Plot	0	0	31												
4	Public Tap/Standpipe	1	0	3	147											
5	Tube-Well/Borehole	0	1	7	47	741										
6	Protected Dug Well	0	0	4	20	82	147									
7	Unprotected Dug Well	0	0	15	23	145	27	248								
8	Protected Spring	0	0	0	1	4	4	1	7							
9	Unprotected Spring	0	0	1	3	26	7	8	0	43						
10	Rainwater Collection	3	4	30	124	623	121	199	6	37	995					
11	Bottled Water	3	1	4	40	86	21	25	1	4	166	182				
12	Cart with Small Tank/Drum	0	0	3	2	3	2	4	0	2	13	3	15			
13	Tanker-Truck	0	0	4	4	7	3	11	0	3	36	12	8	36		
14	Surface Water (River, Dam, Lake, Pond)	3	0	9	73	395	80	105	4	8	583	101	8	11	678	
15	Other	0	0	3	0	0	0	6	0	0	12	3	0	2	3	14

Information from the Community Survey

The objective of the community survey was to obtain information important for interpreting the household survey data. The focus was on the facilities in the community and its general condition. Questions were taken from the GLSS5+ survey, although it asked many more.

Entries in Table 5.40 give a sense of how the place would appear to a visitor. The top three entries concern public facilities. Very few communities have a security post, signs giving directions to public facilities, or even community centers. The last two entries concern the state of commerce: about a quarter of communities have no shops and 20 percent have empty stores.

The sample communities' water sources are catalogued in Table 5.41. As expected there multiple sources in most communities and multiple outlet for the sources a community has. Boreholes are the most common improved facility and the average community having them has three. Boreholes are also the most important water source in 45 percent of communities. This pattern is consistent with the information from the household survey.

Table 5.42 provides additional information on each of the water sources in at least some of the communities. Perhaps the most interesting information is the two sets of columns giving data on the condition of the facilities and the adequacy of the supply of water from the source. The entries show a substantial variance in both dimensions.

Interviewers were instructed to make observations about conditions in the community. The ratings of conditions are summarized in Table 5.43. The statements the interviewers used in making their ratings were consistently from a negative perspective, i.e., the air in the community is generally polluted. They then rated that they agreed or disagreed with the statement. For the first five items in the table, ratings were quite positive with 70-90 percent disagreeing or disagreeing strongly with the negative statement. But even among these in two cases one-quarter of interviewers agreed with the statement—these had to do with garbage and manure piles were evident. In contrast the bottom four conditions listed in the table were rated as problem with 25 to 55 percent of respondents agreeing or strongly agreeing that the conditions existed. The worst score was for the condition of the community square; second was people defecating in the bush.

Table 5.40: Indicators of Community Conditions

Condition	% of communities		
	No	Yes	Not Applicable
Neighborhood security post	93	3	4
Signs/sign boards with public facility directions	90	9	1
Community centre	86	12	2
Shops or buildings that have been damaged or burned	73	5	22
Shops that have closed/gone out of business	50	20	30

The presence of schools in a community likely has a strong influence on education attainment, because with the school close by children can still carry significant responsibilities at home. Mothers' educational attainment is closely related to the well-being of infants and toddlers. Table 5.44 shows that pre-school and primary school coverage in the sample communities is strong: 94 percent have a primary school. There is a significant fall off thereafter: 75 percent have a junior secondary school and just 12 percent a senior secondary school.

Table 5.41: Types of water sources used in communities

Water Source	% of Communities with Water Source	Mean Number of Facilities/Sources for Communities with Water Source	% of Communities for Which This is the Most Important Source
Improved Facilities			
Reservoir/Storage/Water Treatment Plant Pipes Water into Dwelling, Plot, or Yard	12	3.8	4
Public Tap/Standpipe Fed by Reservoir/Storage/Water Treatment Plant	20	3.8	11
Tube Well/Borehole	77	2.9	45
Protected Hand Dug Well	20	3.9	5
Protected Spring	1	3	0
Rainwater Tanks	10	3.7	0
Unimproved Facilities			
Unprotected Hand Dug Well	33	3.3	7
Unprotected Spring	10	3.8	0
Surface Water (River, Dam, Lake, Pond, Stream, Canal, Irrigation Canal)	82	2.7	28
Total	NA	NA	100

Table 5.42: Characteristics of communities' most important water source

Most Important Source	% of Communities for Which This is the Most Important Source	Mean Number in the Community	Condition of Facility (% of communities with this source as the most important)			Water Adequacy (% of communities with this source as the most important)				Who Built the Facility? (% of communities with this source as the most important)				
			Good/Very Good	Not too Good	Poor/Very Poor	Always	Varies With the Season	Varies With Disregard to the Season	Never	District Assembly	NGO	Community	Relig Group	Other
Improved Facilities														
Reservoir/Storage/Water Treatment Plant Pipes Water into Dwelling, Plot, or Yard	4.0	6.0	100.0	0.0	0.0	75.0	25.0	0.0	0.0	25.0	0.0	0.0	0.0	75.0
Public Tap/Standpipe Fed by Reservoir/Storage/Water Treatment Plant	11.0	4.2	90.9	9.1	0.0	45.5	36.4	18.2	0.0	9.1	45.5	18.2	0.0	27.3
Tube Well/Borehole	45.0	2.7	64.4	31.1	4.4	60.0	22.2	11.1	6.7	22.2	68.9	4.4	0.0	4.4
Protected Hand Dug Well	5.0	2.4	60.0	40.0	0.0	60.0	20.0	20.0	0.0	20.0	80.0	0.0	0.0	0.0
Protected Spring	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--
Rainwater Tanks	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--
Unimproved Facilities														
Unprotected Hand Dug Well	7.0	24.4	42.9	42.9	14.3	28.6	57.1	14.3	0.0	0.0	14.3	57.1	0.0	28.6
Unprotected Spring	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--
Surface Water (River, Dam, Lake, Pond, Stream, Canal, Irrigation Canal)	28.0	1.5	17.9	35.7	46.5	42.9	57.1	0.0	0.0	0.0	0.0	10.7	0.0	89.3

Table 5.43: Interviewer's direct observations about the community

Observation	% Distribution				
	Strongly Disagree	Disagree	No Opinion	Agree	Strongly Agree
The air in this community/township is generally polluted.	35	52	1	11	1
There are generally exposed garbage piles visible along the roads.	17	56	2	25	0
There are generally exposed cattle manure piles in this community/township.	18	51	4	22	5
There are blocked drains or obstructed river/gutter in this community/township.	34	52	9	2	3
There are standing pools of water (not including marshes, lakes) in this community/township.	25	64	3	5	3
The house yards in this community/township are generally not swept clean.	13	44	8	30	5
The grass growing in public squares in this community/township are generally not well cared for.	9	32	7	31	21
There are generally many flies visible (near the food vendors).	9	58	8	19	6
People are defecating in the bush	14	26	14	22	24

Table 5.44: Schools in communities, by school type, sponsor, and number

All Communities	
School Type	**% of Community**
Nursery, Kindergarten	83
Primary	94
JSS	75
SSS	12
Other	0
Number of Schools	
1	38
2	24
3-4	19
5-7	10
8-10	9
School Sponsor	**% communities with school built by sponsor**
Government	95
Non-Profit, Private	7
For Profit, Private	26
Other	1

References

Aiga, H., and T. Umenai. 2002. "Impact of Improvement of Water Supply on Household Economy in a Squatter Area of Manila," *Social Science and Medicine*, Vol.55, pp. 627-41.

Cattaneo, M., S. Galliani, P. Gertler, S. Martinez, and R. Titiunik. 2009. "Housing, Health, and Happiness," *American Economic Journal: Economic Policy*, vol.1, no.1, pp.75-105.

Coulombe, H. 2005. "Ghana Census-based Poverty Map: District and Sub-District Results." Accra: Ghana Statistical Service.

Fofack, H. 2000. "Combining Light Monitoring Surveys with Integrated Surveys to Improve Targeting Poverty Reduction: The Case of Ghana," *World Bank Economic Review*, vol.14, no.1, pp.195-219.

Gleick, P.H. 1996. "Basic Water Requirements for Human Activities: Meeting Basic Needs," *Water International*, vol.21, 1996, pp. 83-92.

GSS [Ghana Statistical Service]. 2005. Ghana 2003 Welfare Indicators Questionnaire (CWIQ) Survey Report: Statistical Abstract. Accra: GSS.

Hutton, G., L. Haller, and J. Bartram. 2006. *Economic and Health Effects of Increasing Coverage of Low-cost Water and Sanitation Interventions.* New York: UNDP Human Development Report Office, Occassional Paper 2006/33.

Jalan, J., and M. Ravalion. 2003. "Does Piped Water Reduce Diarrhea for Children in Rural India?" *Journal of Econometrics*, vol.112, pp.153-73.

Kosek, M., C. Bern, and R.L. Guerrant. 2003. "The Global Burden of DiarrhoealDisease, as Estimates from Studies Published between 1992 and 2000, *Bulletin of the World Health Organization*, vol.81, pp. 197-204.

Pattanayak, S. C. Poulos, J.-C. Yang, and S. Patil. 2010. "How Valuable are Environmental Health Interventions? Evaluation of Water and Sanitation Programmes in India," *Bulletin of the World Health Organization*, vol.88, pp.535-42.

Reed, B.J. n.d. "Minimum Water Quantity Needed for Domestic Uses." New Delhi: World Health Organization, Regional Office for Southeast Asia, Technical Note No. 9.

UN-HABITAT. 2003. *The Challenge of Slums: Global Report on Human Settlements,* Sterling Virginia and London: Earthscan Publications

Van Koppen, B., S. Smits, and M. Mikhail. 2009. "Homestead- and Community-Scale Multiple-Use Water Services: Unlocking New Investment Opportunities to Achieve Millenium Challenge Goals," *Irrigation and Drainage*, vol. 58, pp. S73-S86.

Waddington, H., and B. Snilstveit. 2009. "Effectiveness and Sustainability of Water, Sanitation, and Hygiene Interventions in Combating Diarrhoea," *Journal of Development Effectiveness*, vol.1, no.3, pp.295-335.

Annex A: Questionnaires

Water Survey Household Identification Sheet

Location_____ Location ID_____

Interviewer ID: _____ FS ID:_____ Household Number: _____ Coordinates:_____

Attempt	Date	Time	Contact Number	Contact Name & Title	Disp	Notes
1						
2						
3						

Completed (all modules are complete): 01

No Household Member at home or no competent respondent at home at time of interview: 02

Entire household absent for extended period: 03

Postponed/Rescheduled (interview was postponed and a new time scheduled): 04

Final Refusal (interview was refused/no interview completed): 05

Dwelling vacant or address not a dwelling: 06

Other Non-Interview (specify): 09

Partial Complete/Will Return (interview was stopped but will continue later): 10

Partial Complete/Interview Finished (interview was stopped and will not continue): 11

Temporary Refusal (interview was refused, FS will follow-up): 12

Out of Scope (the household is not within the sample): 13

MiDA Ghana Water Household Questionnaire v10

1. Date of Interview (DD/MM/YYYY): |_|_|/|_|_|/|_|_|_|_|

2. Locality: _____ Code: |_|_|

3. Household No. |_|_|

4. Survey Start Time:_____ Survey End Time:_____

5. Is this household in a structure? (0 No, 1 Yes) |_|

6. GIS Location of Household:
Latitude: N/S |_| Degrees |_|_| Minutes |_|_|.|_|_|_|
Longitude: E/W |_| Degrees |_|_|_| Minutes |_|_|.|_|_|_|

7. Interviewer ID	Supervisor ID	Editor ID	Data Enterer ID
|_|_|_|_|	|_|_|_|_|	|_|_|_|_|	|_|_|_|_|

8. Interviewer/supervisor comments:

Introduction and Consent

Hello and thank you for talking to us. We are from Panafields survey firm. We are working on a Water Project being implemented by the Government of Ghana, and we would like to interview you so that we can understand how improved water services affect communities like yours.

The interview will take about 90 minutes and we will ask questions to you or to other adults in the household who may be the most knowledgeable about particular topics. This survey will be repeated in 2011 and you may be asked to participate again so that we can measure the changes.

All of the information you and others in your household give us will remain confidential and you will not be identified. Results will not be presented for individuals or households. So please feel free to speak openly. The information that you provide will be kept until 2012 for the purposes of preparing reports to the project sponsor. Your participation is completely voluntary. You are free to not answer any question with which you are not comfortable, and you may stop the interview at any time.

If you have any questions about the survey, you may contact [CONTACT PERSON NAME] at Panafields at the telephone number or address on this card.
GIVE CARD TO RESPONDENT

Do you wish to participate in this survey? May we start now?

IF YOU ADMINISTER PART OF THE QUESTIONNAIRE TO OTHER MEMBERS OF THE HOUSEHOLD (AS IS LIKELY), YOU MUST READ THIS GREETING/CONSENT TO EACH PERSON WHO PARTICIPATES IN THE SURVEY AND RECORD THEIR ID NUMBER (FROM SECTION B, PAGE 3) AND RESPONSE TO CONSENT BELOW.

Member ID of Respondent #1 |_|_| Response to consent (*0 NO, 1 YES*) |_|
Member ID of Respondent #2 |_|_| Response to consent (*0 NO, 1 YES*) |_|
Member ID of Respondent #3 |_|_| Response to consent (*0 NO, 1 YES*) |_|

May I please talk to the person who knows most about your household's acquisition and use of water and economic activities? *ASK THE PERSON THE MOST KNOWLEDGEABLE ABOUT WATER USE TO RESPOND TO SECTION 5 AND 6, AND THE PERSON MOST KNOWLEDGEABLE ABOUT ECONOMIC ACTIVITIES TO RESPOND TO SECTION 4.*

Respondent 1

a. What is your name?

_____

b. What is your family name?

_____

Sections Completed:

Respondent 2

a. What is your name?

_____

b. What is your family name?

_____

Sections Completed:

9. Main language spoken at home: |_|_| Specify other _____

11	Asante	21	Ga	53	Frafra/Gruni
12	Fanti	22	Dangme	54	Kassene
13	Akuapem	31	Ewe	55	Dagbani
14	Sefwi	41	Guan	56	Wali/Dagari
15	Brong	51	Buli	57	Sissala
16	Nzema	52	Mamprusi	96	Other (specify)

10. Type of housing: |_| Specify other _____

1	Separate houses (bungalow)	4	Room(s) [compound]	7	Several huts/buildings (different compounds)
2	Semi-detached house	5	Room(s)	8	Tents/improvised home
3	Flat/apartment	6	Several huts/buildings (same compounds)	9	Other (specify)

11. Occupancy status |_| Specify other _____

1	Own	4	Perching
2	Rent	5	Other (specify)
3	Rent-free		

12. Does another household share this dwelling? 0 No, 1 Yes |_|

SECTION 1: HOUSEHOLD MEMBERS

Member ID	1.1 Please tell me the name of each of the members of this household, starting with yourself?	1.2 What is [NAME]'s sex? 1 Male 2 Female	1.3 What is [NAME]'s relationship with the head of household? SEE CODES	1.5 How old is [NAME]? YEARS AND MONTHS *IF <5 YEARS OLD* ➔*1.10* *IF 5-11 YEARS OLD* ➔ *1.14* Years	1.6 Months	1.7 *(IF 1.5>11 years)* What is [NAME]'s current marital status? 1 Married 2 Consensual union 3 Separated ➔*1.14* 4 Divorced ➔*1.14* 5 Widowed ➔*1.14* 6 Never married ➔*1.14*	1.8 *(IF 1.5>11 years)* Does [NAME]'s spouse live in this household? 0 No ➔*1.14* 1 Yes	1.9 *(IF 1.5>11 years)* If spouse is in the household, enter spouse ID. IF MORE THAN ONE SPOUSE, ENTER THE FIRST ONE ➔*1.14*
01		_|	__|	__|	__|	_|	_|	__|
02		_|	__|	__|	__|	_|	_|	__|
03		_|	__|	__|	__|	_|	_|	__|
04		_|	__|	__|	__|	_|	_|	__|
05		_|	__|	__|	__|	_|	_|	__|
06		_|	__|	__|	__|	_|	_|	__|
07		_|	__|	__|	__|	_|	_|	__|
08		_|	__|	__|	__|	_|	_|	__|
09		_|	__|	__|	__|	_|	_|	__|
10		_|	__|	__|	__|	_|	_|	__|
11		_|	__|	__|	__|	_|	_|	__|
12		_|	__|	__|	__|	_|	_|	__|
13		_|	__|	__|	__|	_|	_|	__|
14		_|	__|	__|	__|	_|	_|	__|
15		_|	__|	__|	__|	_|	_|	__|
16		_|	__|	__|	__|	_|	_|	__|
17		_|	__|	__|	__|	_|	_|	__|

Codes for 1.3

01 Head
02 Spouse (wife/husband)
03 Child (son/daughter)
04 Grandchild
05 Parent/parent-in-law
06 Son-/daughter-in-law
07 Other relative
08 Adopted/ foster/stepchild
09 Househelp
10 Non-relative

18		|_|	|_|_|	|_|_|	|_|_|	|_|	|_|	|_|_|	

Member ID	1.10	1.11	1.12	1.13
	(IF 1.5<5) Do you have a weight card for [NAME]? 0 No ➔ ***Go to 1.14*** 1 Yes	*(IF 1.5<5)* IF 1.10=1, GET THIS INFORMATION FROM WEIGHT CARD What is the last time [NAME] was weighed?		*(IF 1.5<5)* IF 1.10=1, GET THIS INFORMATION FROM WEIGHT CARD In how many of the last 6 months was [NAME] weighed? 0-6 MONTHS
		MONTH	YEAR	
01	|_|	|_|_|	|_|_|_|_|	|_|
02	|_|	|_|_|	|_|_|_|_|	|_|
03	|_|	|_|_|	|_|_|_|_|	|_|
04	|_|	|_|_|	|_|_|_|_|	|_|
05	|_|	|_|_|	|_|_|_|_|	|_|
06	|_|	|_|_|	|_|_|_|_|	|_|
07	|_|	|_|_|	|_|_|_|_|	|_|
08	|_|	|_|_|	|_|_|_|_|	|_|
09	|_|	|_|_|	|_|_|_|_|	|_|
10	|_|	|_|_|	|_|_|_|_|	|_|
11	|_|	|_|_|	|_|_|_|_|	|_|
12	|_|	|_|_|	|_|_|_|_|	|_|
13	|_|	|_|_|	|_|_|_|_|	|_|
14	|_|	|_|_|	|_|_|_|_|	|_|
15	|_|	|_|_|	|_|_|_|_|	|_|
16	|_|	|_|_|	|_|_|_|_|	|_|
17	|_|	|_|_|	|_|_|_|_|	|_|
18	|_|	|_|_|	|_|_|_|_|	|_|

Member ID	1.14 Does [NAME]'s mother live in this household? 0 No ➔*1.16* 1 Yes	1.15 ID number of mother	1.16 What was [NAME]'s mother's highest educational level completed? SEE CODES	1.17 What kind of work has [NAME]'s mother done for most of her life? SEE CODES	1.18 *(IF 1.17=10)* Specify other.
01	_|	_|_|	_|_|	_|_|	
02	_|	_|_|	_|_|	_|_|	
03	_|	_|_|	_|_|	_|_|	
04	_|	_|_|	_|_|	_|_|	
05	_|	_|_|	_|_|	_|_|	
06	_|	_|_|	_|_|	_|_|	
07	_|	_|_|	_|_|	_|_|	
08	_|	_|_|	_|_|	_|_|	
09	_|	_|_|	_|_|	_|_|	
10	_|	_|_|	_|_|	_|_|	
11	_|	_|_|	_|_|	_|_|	
12	_|	_|_|	_|_|	_|_|	
13	_|	_|_|	_|_|	_|_|	
14	_|	_|_|	_|_|	_|_|	
15	_|	_|_|	_|_|	_|_|	
16	_|	_|_|	_|_|	_|_|	
17	_|	_|_|	_|_|	_|_|	
18	_|	_|_|	_|_|	_|_|	

Codes for 1.16

01 None
02 Primary
03 Middle/JSS
04 Voc./Comm.
05 O' Level
06 SSS
07 A' Level
08 Training college
09 Tech./Prof.
10 Tertiary
11 Basic Religious Education

Codes for 1.17

01 Professional/technical
02 Administrative/managerial
03 Clerical
04 Sales
05 Service
06 Agric./Animal Husbandry/Forestry/Fishing/Hunting
07 Production & related workers
08 Workers NEC
09 Homemaker
10 Other (specify)

SECTION 2: HOUSEHOLD EDUCATION

Fill out for all household members 3 years and older.

Member ID	2.1 Has [NAME] ever attended school? 0 No ➔*NEXT MEMBER* 1 Yes	2.2 What was the highest grade completed? SEE CODES	2.3 *(IF 2.2=61)* Specify other.	2.4 What was the highest educational qualification attained? SEE CODES
01	|_|	|_|_|		|_|_|
02	|_|	|_|_|		|_|_|
03	|_|	|_|_|		|_|_|
04	|_|	|_|_|		|_|_|
05	|_|	|_|_|		|_|_|
06	|_|	|_|_|		|_|_|
07	|_|	|_|_|		|_|_|
08	|_|	|_|_|		|_|_|
09	|_|	|_|_|		|_|_|
10	|_|	|_|_|		|_|_|
11	|_|	|_|_|		|_|_|
12	|_|	|_|_|		|_|_|
13	|_|	|_|_|		|_|_|
14	|_|	|_|_|		|_|_|
15	|_|	|_|_|		|_|_|
16	|_|	|_|_|		|_|_|
17	|_|	|_|_|		|_|_|
18	|_|	|_|_|		|_|_|

Codes for 2.2

00 None
01 Pre-school
11 P1
12 P2
13 P3
14 P4
15 P5
16 P6
17 JHS1
18 JHS2
19 JHS3
20 M1
21 M2
22 M3
23 M4
24 SHS1
25 SHS2
26 SHS3
27 S1
28 S2
29 S3
30 S4
31 S5
32 L6
33 U6
41 Voc/Tech/Computer/ Comm/Agric
42 Teacher training
43 Nursing
51 Polytechnic
52 University
53 Other Tertiary
61 Other (specify)

Codes for 2.4

01 None
02 MSLC
03 BECE
04 Voc./Comm.
05 Teacher Training Cert. A
06 Teacher Training Post Sec
07 GCE 'O' Level
08 SSCE
09 GCE 'A' Level
10 Tech./Prof. Cert.
11 Tech./Prof. Dip.
12 HND
13 Bachelors
14 Masters
15 Doctorate
16 Other

SECTION 3: HOUSEHOLD HEALTH IN THE LAST TWO WEEKS

Fill out for all household members.

Member ID	3.1 During the last two weeks, has [NAME] suffered from either illness or injury? 0 Neither ➔*NEXT MEMBER* 1 Injury Only ➔*NEXT MEMBER* 2 Illness Only 3 Both	3.2 What was the illness that [NAME] suffered? SEPARATE MULTIPLE RESPONSES WITH COMMAS SEE CODES **IF 3.1 = 0 or 1 GO TO 4.1** **IF 3.2=5** ➔*NEXT MEMBER*	3.3 What symptoms did [NAME] experience? 1 Abdominal Pain 2 Frequent Evacuation 3 Watery Feces 4 Bloody Feces 5 Vomiting 6 Fever SEPARATE MULTIPLE RESPONSES WITH COMMAS	3.4 What treatments did [NAME] receive? SEPARATE MULTIPLE RESPONSES WITH COMMAS SEE CODES	3.5 For how many days during the last 2 weeks has [NAME] suffered from this condition? 1-14 DAYS	**Codes for 3.2** 1 Diarrhea with blood 2 Watery diarrhea 3 Guinea worm 4 Bilharzia 5 Other **Codes for 3.4** 01 Boiled water 02 No feeding (apart from breastfeeding) 03 Fed easily digestible food e.g. starch in the form of rice-water, etc. 04 Breastfeeding 05 Home-made Oral Rehydration Therapy (ORT) fluids using safe water (boiled) 06 Oral Rehydration Therapy (ORT) sachets using safe water (boiled) 07 Less water than usual 08 Zinc supplement 09 Traditional remedies such as cold tea and pumpkin seeds 10 Other traditional remedy 11 Medicines from clinic/doctor (other than those above) 12 Extra water without boiling 13 No treatment
01	_|				_|_|	
02	_|				_|_|	
03	_|				_|_|	
04	_|				_|_|	
05	_|				_|_|	
06	_|				_|_|	
07	_|				_|_|	
08	_|				_|_|	
09	_|				_|_|	
10	_|				_|_|	
11	_|				_|_|	
12	_|				_|_|	
13	_|				_|_|	
14	_|				_|_|	
15	_|				_|_|	
16	_|				_|_|	
17	_|				_|_|	
18	_|				_|_|	

Member ID	3.6 During the last 2 weeks, did [NAME] have to stop the usual activities because of this condition? 0 No ➔*3.8* 1 Yes	3.7 For how many days? 1-14 DAYS	3.8 During the last 2 weeks, did [NAME] need to be cared for at home because of this condition? 0 No ➔*3.10* 1 Yes	3.9 For how many days? 1-14 DAYS	3.10 During the last 2 weeks, has [NAME] visited any health facility because of this condition? 0 No ➔*3.12* 1 Yes	3.11 During the last two weeks, how many days did [NAME] receive care at a health facility? 1-14 DAYS	3.12 During the last 2 weeks, was [NAME] admitted to a hospital or health centre on account of this condition? INCLUDE TRADITIONAL HEALING CENTRES 0 No ➔*NEXT MEMBER* 1 Yes	3.13 How many nights did [NAME] stay in the hospital/health centre during the last 2 weeks? 1-14 NIGHTS
01	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
02	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
03	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
04	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
05	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
06	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
07	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
08	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
09	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
10	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
11	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
12	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
13	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
14	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
15	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
16	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
17	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|
18	_|	_|_|	_|	_|_|	_|	_|_|	_|	_|_|

SECTION 4: ACTIVITY STATUS, PRIMARY AND SECONDARY OCCUPATIONS – LAST 12 MONTHS

Fill out for all household members 7 years and older.

Member ID	4.1 Did [NAME] do any work for pay, profit, family gain, or did [NAME] produce anything for barter or home use during the past 12 months? 0 No ➔*NEXT MEMBER* 1 Yes	4.2 During the past 12 months, how many jobs did [NAME] do altogether?	4.3 How many weeks per year and hours per week did [NAME] work over the past 12 months? Wks/Yr	4.4 Hrs/Wk	4.5 What was the status of [NAME] in their MAIN OCCUPATION of the past 12 months? SEE CODES ➔*IF 4.5≠5, GO TO 4.8*	4.6 *(IF 4.5=5)* For whom did [NAME] work in their MAIN OCCUPATION? SEE CODES	4.7 *(IF 4.6=10)* Specify other.
01	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
02	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
03	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
04	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
05	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
06	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
07	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
08	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
09	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
10	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
11	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
12	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
13	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
14	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
15	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	
16	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|	

Codes for 4.5

1 Any kind of business, big or small, for himself/ herself. [15]
2 Help unpaid in a family business of any kind. [16]
3 Help on the family plot, food garden, cattle post.[17]
4 Collect vegetables or catch animals or other food for sale or as family food.
5 Do any work for a wage, salary, or any payment in kind, even if only for one hour.
6 Beg for money or food in public.

Codes for 4.6

Government Sector:
01 Civil service
02 Other public service
03 Parastatals
04 NGOs
05 Cooperatives
06 Inter. Organ./ Diplomatic Mission
07 Private sector formal (incl. paid apprentices)
08 Private sector informal
09 Agric. Business
10 Other (specify)

[15] (Examples: Selling things, making things for sale, repairing things, guarding cars, donkey cart or other transport business etc.)
[16] (Examples: Help to sell or make things for sale, cleaning up for a business, etc. Don't count normal housework)
[17] (Examples: ploughing, harvesting, weeding, looking after livestock)

17	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|		
18	|_|	|_|_|	|_|_|	|_|_|	|_|	|_|_|		

Member ID	4.8	4.9	4.10	4.11	4.12	4.13	4.14	4.15	4.16
	Has [NAME] received or will [NAME] receive money for work in their MAIN OCCUPATION? 0 No ➔*4.13* 1 Yes	What was the amount (incl. any bonuses, commissions allowances, or tips) received? TIME UNIT CODES 1 Daily 2 Weekly 3 Fortnightly (every two weeks) 4 Monthly 5 Quarterly (every 3 months) 6 Yearly			Are taxes already deducted from [NAME]'s pay for their MAIN OCCUPATION? 0 No 1 Yes	Does [NAME] receive any payment for work in their MAIN OCCUPATION in the form of goods and services? 0 No ➔*4.17* 1 Yes	What is the value of goods or services provided? TIME UNIT CODES 1 Daily 2 Weekly 3 Fortnightly (every two weeks) 4 Monthly 5 Quarterly (every 3 months) 6 Yearly		
		GH¢	P	TIME UNIT			GH¢	P	TIME UNIT
01	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
02	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
03	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
04	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
05	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
06	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
07	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
08	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
09	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
10	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
11	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
12	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
13	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
14	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
15	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
16	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
17	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|
18	|_|	|_|_|_|_|	|_|_|	|_|	|_|	|_|	|_|_|_|_|	|_|_|	|_|

Member ID	4.17 During the past 12 months, for how many weeks did [NAME] do work for their MAIN OCCUPATION? WEEKS	4.18 During these weeks, how many hours per week did [NAME] usually work in their MAIN OCCUPATION? HOURS/WEEK	4.19 During the past 12 months, did [NAME] do any work beside the MAIN OCCUPATION? 0 No ➔*NEXT MEMBER* 1 Yes	4.20 What was the status of [NAME] in their SECONDARY OCCUPATION of the past 12 months? SEE CODES ➔*IF 4.20≠5, GO TO 4.23*	4.21 *(IF 4.20=5)* For whom did [NAME] work in their SECONDARY OCCUPATION? SEE CODES	4.22 *(IF 4.21=10)* Specify other.
01	__	___	_	_	__	
02	__	___	_	_	__	
03	__	___	_	_	__	
04	__	___	_	_	__	
05	__	___	_	_	__	
06	__	___	_	_	__	
07	__	___	_	_	__	
08	__	___	_	_	__	
09	__	___	_	_	__	
10	__	___	_	_	__	
11	__	___	_	_	__	
12	__	___	_	_	__	
13	__	___	_	_	__	
14	__	___	_	_	__	
15	__	___	_	_	__	
16	__	___	_	_	__	
17	__	___	_	_	__	
18	__	___	_	_	__	

Codes for 4.20

1 Any kind of business, big or small, for himself/ herself. [18]
2 Help unpaid in a family business of any kind. [19]
3 Help on the family plot, food garden, cattle post. [20]
4 Collect vegetables or catch animals or other food for sale or as family food.
5 Do any work for a wage, salary, or any payment in kind, even if only for one hour.
6 Beg for money or food in public.

Codes for 4.21

Government Sector:
01 Civil service
02 Other public service
03 Parastatals
04 NGOs
05 Cooperatives
06 Inter. Organ./ Diplomatic Mission
07 Private sector formal (incl. paid apprentices)
08 Private sector informal
09 Agric. Business
10 Other (specify)

[18] (Examples: Selling things, making things for sale, repairing things, guarding cars, donkey cart or other transport business etc.)
[19] (Examples: Help to sell or make things for sale, cleaning up for a business, etc. Don't count normal housework)
[20] (Examples: ploughing, harvesting, weeding, looking after livestock)

Member ID	4.23	4.24	4.25	4.26	4.27	4.28	4.29	4.30	4.31
	Has [NAME] received or will [NAME] receive money for work in their SECONDARY OCCUPATION? 0 No ➔*4.28* 1 Yes	What is the amount (incl. any bonuses, commissions, allowances, or tips) received for work in the SECONDARY OCCUPATION? TIME UNIT CODES 1 Daily 2 Weekly 3 Fortnightly (every two weeks) 4 Monthly 5 Quarterly (every 3 months) 6 Yearly			Are taxes already deducted from [NAME]'s pay for the SECONDARY OCCUPATION? 0 No 1 Yes	Does [NAME] receive any payment for work in their SECONDARY OCCUPATION the form of goods and services? 0 No ➔ *Section 5* 1 Yes	What is the value of goods or services provided? TIME UNIT CODES 1 Daily 2 Weekly 3 Fortnightly (every two weeks) 4 Monthly 5 Quarterly (every 3 months) 6 Yearly		
		GH¢	P	TIME UNIT			GH¢	P	TIME UNIT
01	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
02	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
03	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
04	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
05	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
06	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
07	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
08	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
09	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
10	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
11	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
12	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
13	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
14	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
15	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
16	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
17	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|
18	_|	_|_|_|_|	_|_|	_|	_|	_|	_|_|_|_|	_|_|	_|

Member ID	4.32	4.33	4.34	4.35
	During the past 12 months, for how many weeks did [NAME] do work for their SECONDARY OCCUPATION?	During these weeks, how many hours per week did [NAME] usually work in their SECONDARY OCCUPATION?	Did [NAME] work on this job at the same time as his/her main job? 0 No ➔*NEXT MEMBER* 1 Yes	How many weeks did [NAME] do both together? ENTER 00 FOR <1 WEEK ➔*NEXT MEMBER*
	WEEKS	HOURS/WEEK		
01	__	___	_	__
02	__	___	_	__
03	__	___	_	__
04	__	___	_	__
05	__	___	_	__
06	__	___	_	__
07	__	___	_	__
08	__	___	_	__
09	__	___	_	__
10	__	___	_	__
11	__	___	_	__
12	__	___	_	__
13	__	___	_	__
14	__	___	_	__
15	__	___	_	__
16	__	___	_	__
17	__	___	_	__
18	__	___	_	__

SECTION 5: WATER USE AND SOURCES – Ask questions 5.1-5.20 before moving on to the next source; after going through all sources, go to 5.21

5.1			5.2	5.3	5.4	5.5	5.6	5.7	5.8
What are the sources of water used by your household?[21] 0 Not used by household ➔*NEXT SOURCE* 1 Used by household			When do you mostly use water from this source? 1 Dry season only (when there is water scarcity) 2 Wet/rainy season only ➔*NEXT SOURCE* 3 All seasons	During the dry season, for what purpose do you mostly use water from this source? SEE CODES SEPARATE MULTIPLE RESPONSES WITH COMMAS	*(IF 5.3=5 or 6)* Specify.	During the dry season, Is water from this source safe to drink? 0 No 1 Yes ➔*5.9*	During the dry season, how often do you treat water from this source for drinking? 1 Always 2 Most of the time 3 Sometimes 4 Never ➔*5.9*	During the dry season, what treatment methods do you use? SEE CODES	*(IF 5.7=7)* Specify other.
1	Piped water into dwelling	|_|	|_|			|_|	|_|	|_|	
2	Piped water to yard/plot	|_|	|_|			|_|	|_|	|_|	
3	Piped into someone else's yard/plot	|_|	|_|			|_|	|_|	|_|	
4	Public tap/standpipe	|_|	|_|			|_|	|_|	|_|	
5	Tube-well/borehole	|_|	|_|			|_|	|_|	|_|	
6	Protected dug well	|_|	|_|			|_|	|_|	|_|	

Codes for 5.3
0 None
1 Drinking and cooking
2 Cleaning the house
3 Washing and taking baths
4 Provision for animals
5 Economic use (specify)
6 Other (specify)

Codes for 5.7
1 Boiling
2 Add bleach/ chlorine
3 Strain through cloth
4 Use water filter (ceramic, sand, composite, etc)
5 Let it stand and settle
6 Solar

[21] Interviewers will have sketches of all types of water sources to show the respondent to ensure that there is a common understanding of the terminology being used.

7	Unprotected dug well	_|	_|			_|	_|	_|		disinfection 7 Other (specify)
8	Protected spring	_|	_|			_|	_|	_|		
9	Unprotected spring	_|	_|			_|	_|	_|		
10	Rainwater collection	_|	_|			_|	_|	_|		
11	Bottled water	_|	_|			_|	_|	_|		
12	Cart with small tank/drum	_|	_|			_|	_|	_|		
13	Tanker-truck	_|	_|			_|	_|	_|		
14	Surface water (river, dam, lake, pond)	_|	_|			_|	_|	_|		
15	Other (specify) _____ _____ ____	_|	_|			_|	_|	_|		

If non-household members fetch water enter 0 at 5.9 and Go to 5.14

Water Source		5.9		5.10	5.11	5.12	5.13
		During the dry season, who collects water from this source most of the time?		For EACH TRIP during the dry season, what is the average time (in minutes) spent:			During the dry season, how many times per week does this member collect water from this source?
				Travelling to and from the water source	Waiting in line at the water source	Filling containers with water	
1	Piped water into dwelling	Person 1	_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|
		Person 2	_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|
2	Piped water to yard/plot	Person 1	_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|
		Person 2	_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|_|	_|_|_|

Water Source		5.9		5.10	5.11	5.12	5.13
		During the dry season, who collects water from this source most of the time?		For EACH TRIP during the dry season, what is the average time (in minutes) spent:			During the dry season, how many times per week does this member collect water from this source?
				Travelling to and from the water source	Waiting in line at the water source	Filling containers with water	
3	Piped into someone else's yard/plot	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
4	Public tap/standpipe	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
5	Tube-well/borehole	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
6	Protected dug well	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
7	Unprotected dug well	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
8	Protected spring	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
9	Unprotected spring	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
10	Rainwater collection	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
11	Bottled water	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
12	Cart with small tank/drum	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
13	Tanker-truck	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
14	Surface water (river, dam, lake, pond)	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
15	Other (from 5.1)	Person 1	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_
		Person 2	_|_|_	_|_|_|_	_|_|_|_	_|_|_|_	_|_|_

			5.14	5.15	5.16	5.17	5.18	5.19	5.20
For each person who fetches water from each source please show me the container they use most of the time. **MEASURE THE CONTAINER**			Width of container from center in centimeters.	Height of container in centimeters	If container has volume printed on it on then enter the volume here.	Please enter units used for 5.16 1 Liters 2 Gallons	During the dry season, how many containers of this type are usually filled in ONE TRIP?	During the dry season, do you pay for water from this source by month or by quantity of water collected? 0 Do not pay ➔ *NEXT MEMBER/SOURCE* 1 per month 2 per container	During the dry season, how much do you pay per month or per container? ➔ *NEXT MEMBER/ SOURCE*
1	Piped water into dwelling	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
2	Piped water to yard/plot	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
3	Piped into someone else's yard/plot	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
4	Public tap/standpipe	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
5	Tube-well/borehole	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
6	Protected dug well	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
7	Unprotected dug well	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
8	Protected spring	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
9	Unprotected spring	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
10	Rainwater collection	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
11	Bottled water	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
12	Cart with small tank/drum	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
		Person 2	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|
13	Tanker-truck	Person 1	_|_|_|	_|_|_|		_|	_|_|_|	_|_|_|	_|_|_|_| / _|_|

		Person 2	___	___		_	___	___	____ / ___
14	Surface water	Person 1	___	___		_	___	___	____ / ___
		Person 2	___	___		_	___	___	____ / ___
15	Other (from 5.1)	Person 1	___	___		_	___	___	____ / ___
		Person 2	___	___		_	___	___	____ / ___

5.21	Did your household sell any water to someone else during the dry season?	0 No ➔*5.23* 1 Yes	_
5.22	During the dry season, how much per week did your household receive for the water sold?	GH¢	____
		P	___
5.23	If the household does not fetch its water but rather has a connection in the house or so close in the yard that it is not really "fetching" the water: How much water do you think you use a day normally?	Unit (1=liters, 2=gallons)	_
5.24		Quantity	____

SECTION 6: TOILETS AND HYGIENE

6.1	What type of toilet is used by your household?	**Codes** Flush/pour to… 1 …piped sewer system 2 …septic tank 3 …pit latrine 4 …elsewhere 5 …don't know 6 Ventilated improved pit latrine 7 Pit latrine with slab 8 Pit latrine without slab (open pit) 9 Pan/bucket 10 VIP/KVIP, public facility 11 No facilities	__
6.2	How was this built and paid for?	1 Owner built, family savings and support 2 Owner built, subsidy from govt or NGO 3 Govt/NGO built 4 Other (specify)	_
6.3	*(IF 6.2=4)* Specify other.		

6.4	Does your household have the following hygiene facilities?	Handwashing facility, special basin	_	**Codes for 6.9** 1 Child used toilet/latrine 2 Put/rinsed into toilet or latrine
6.5	0 No 1 Yes	Soak-pit (soak-away pit)	_	
6.6		Rubbish pit	_	
6.7	Did your household participate in hygiene promotion in the past year? 0 No 1 Yes		_	

6.8	Have there been visits to the village to follow up on the hygiene promotion? 0 No 1 Yes	|_|	3 Put/rinsed into drain or soak-away pit 4 Thrown into garbage 5 Buried 6 Left in the open 7 Other (specify)
6.9	The last time the youngest child passed stools, how were these disposed of?	|_|	
6.10	*(IF 6.9=7)* Specify other.		

Questions on Personal Hygiene:

6.11	How often do you feel you need to wash your hands with soap after using a toilet?	|_|	**Codes** 1 Always 2 Sometimes 3 Never
6.12	How often do you feel you need to wash your hands with soap before eating?	|_|	
6.13	How often do you feel you need to wash your hands with soap before preparing food?	|_|	

SECTION 7: HOUSEHOLD UNIT, DURABLES, FUEL, POVERTY PREDICTORS

Household Unit:

7.1	What is the main construction material used for the outer wall of the main building?	**Codes** 01 Mud/mud bricks 02 Wood/bamboo 03 Metal sheet/slate/asbestos 04 Stone 05 Burned bricks 06 Cement/sandcrete blocks 07 Landcrete 08 Thatch 09 Cardboard 10 Other (specify)	|_|_|
7.2	*(IF 7.1=10)* Specify other.		
7.3	What is the main construction material used for the floor?	1 Earth/mud/mud bricks 2 Wood 3 Stone 4 Cement/concrete 5 Burnt bricks 6 Vinyl tiles 7 Ceramic/marble/tiles 8 Terrazzo 9 Other (specify)	|_|
7.4	*(IF 7.3=9)* Specify other.		
7.5	What is the main material used for the roof?	1 Palm leaves/raffia/thatch 2 Wood 3 Corrugated iron sheets 4 Cement/concrete 5 Asbestos/slate 6 Roofing tiles 7 Mud bricks/earth 8 Bamboo 9 Other (specify)	|_|
7.6	*(IF 7.5=9)* Specify other.		
7.7	How many rooms does this household occupy? COUNT LIVING ROOMS, DINING ROOMS, BUT NOT BATHROOMS AND KITCHENS		|_|_|

ables:

7.8	Does the household own any of the following? INCLUDE ITEMS ONLY IF THEY ARE IN WORKING CONDITION	0 No 1 Yes
a)	Electric iron	|_|
b)	Refrigerator	|_|
c)	Television	|_|
d)	Video deck	|_|
e)	Cassette player/radio	|_|
f)	Sewing machine	|_|
g)	Bicycle	|_|

Fuel:

7.9	What is the main fuel used for cooking?	**Codes**		|_|
		1 Firewood	5 Electricity	
		2 Charcoal	6 Crop residue/sawdust	
		3 Kerosene/oil	7 Animal waste	
		4 Gas	8 Other, specify	
7.10	*(IF 7.9=8)* Specify other.			
7.11	What is the main fuel used for lighting?	1 Kerosene/paraffin	6 Candles	|_|
		2 Gas	7 Firewood	
		3 Electricity	8 Solar energy	
		4 Generator	9 Other, specify	
		5 Battery		
7.12	*(IF 7.11=9)* Specify other.			

Poverty Predictors:

7.13	How much was spent on soap and washing powder in the last 4 weeks?	GH¢	|_|_|_|_|
7.14		P	|_|_|
7.15	How much was spent on rice in the last 4 weeks?	GH¢	|_|_|_|_|
7.16		P	|_|_|
7.17	How much was spent on bread in the last 4 weeks?	GH¢	|_|_|_|_|
7.18		P	|_|_|
7.19	How much was spent on kenkey in the last 4 weeks?	GH¢	|_|_|_|_|
7.20		P	|_|_|
7.21	How much was spent on tomatoes in the last 4 weeks?	GH¢	|_|_|_|_|
7.22		P	|_|_|

MiDA Ghana Water Community Survey v11

1. Date of Interview (DD/MM/YYYY): |_|_|/|_|_|/|_|_|_|_| Supervisor ID:_____

2. Locality: _____ Code: |_|_|_|_|

3. GIS Location:
Latitude: N/S |_| Degrees |_|_| Minutes |_|_|.|_|_|_|
Longitude: E/W |_| Degrees |_|_|_| Minutes |_|_|.|_|_|_|

4. Interviewer/supervisor comments:

Introduction and Consent

Hello and thank you for talking to us. We are from Panafields survey firm. We are working on a Water Project being implemented by the Government of Ghana, and we would like to interview you so that we can understand how improved water services affect communities like yours.

The interview will take about 15 minutes and we will ask you questions about conditions in your community. This survey will be repeated in 2011 and you may be asked to participate again so that we can measure the changes.

All of the information you give us will remain confidential and you will not be identified. So please feel free to speak openly. The information that you provide will be kept until 2012 for the purposes of preparing reports to the project sponsor. Your participation is completely voluntary. You are free to not answer any question with which you are not comfortable, and you may stop the interview at any time.

If you have any questions about the survey, you may contact [CONTACT PERSON NAME] at Panafields at the telephone number or address on this card.
GIVE CARD TO RESPONDENT

Do you wish to participate in this survey? May we start now?
IF YOU ADMINISTER PART OF THE QUESTIONNAIRE TO MORE THAN ONE PERSON, YOU MUST READ THIS GREETING/CONSENT TO EACH PERSON WHO PARTICIPATES IN THE SURVEY AND RECORD THEIR NAME AND RESPONSE TO CONSENT BELOW.

Title of Respondent #1		Response to consent (*0 NO, 1 YES*)	|_|
Title of Respondent #2		Response to consent (*0 NO, 1 YES*)	|_|
Title of Respondent #3		Response to consent (*0 NO, 1 YES*)	|_|

THESE QUESTIONS ARE ASKED TO THE PERSON WHO JUST AGREED TO BE A RESPONDENT.

Section 1: Screening

If treatment village:			
1.1	Was the water system in this village improved in the past 3-4 months? For example was a new borehole or piping system installed or improved?	0 No ➔*SECTION 2* 1 Yes ➔ *Ask 1.2*	|_|
1.2	Was this improvement part of a MiDA project?	0 No ➔*MUST SUBSTITUTE VILLAGE* 1 It was MiDA ➔ *ask 1.3*	
1.3	Can people now use the improved facility?	0 No -> SECTION 2 1 Yes *If 1.2 = 'YES' YOU MUST SUBSTITUTE VILLAGE*	|_|
If control village:			
1.3	Is this village's water system scheduled for improvement within the next year? **If answer is 1 or 2, stop interview.**	1 Definitely ➔*MUST SUBSTITUTE VILLAGE* 2 Very likely ➔*MUST SUBSTITUTE VILLAGE* 3 Unlikely 4 Definitely not	|_|

Section 2: Public Water Facilities

Facility Type		2.1	2.2	2.3	2.4	2.5	2.6	2.7
		Is there [FACILITY] in this community? 0 No ➔*NEXT FACILITY* 1 Yes	Which facility serves the most households in the community ? MARK ONLY ONE WITH A "1"	How many are in the community ?	What is the condition of facility? 1 Very good 2 Good 3 Not too good 4 Poor 5 Very poor	Is the facility providing an adequate supply of water? 1 Always 2 Varies with the season 3 Varies with disregard to season 4 Never	Who built the facility? 1 District assembly 2 NGO 3 Community 4 Religious group 5 Other (specify)	*(IF 2.6=5)* Specify other.
IMPROVED								
a	Reservoir/storage/water treatment plant pipes water into dwelling, plot, or yard	|_|	|_|	|_|_|	|_|	|_|	|_|	
b	Public tap/standpipe fed by reservoir/storage/water treatment plant	|_|	|_|	|_|_|	|_|	|_|	|_|	
c	Tube well/borehole	|_|	|_|	|_|_|	|_|	|_|	|_|	
d	Protected hand dug well	|_|	|_|	|_|_|	|_|	|_|	|_|	
e	Protected spring	|_|	|_|	|_|_|	|_|	|_|	|_|	
f	Rainwater tanks	|_|	|_|	|_|_|	|_|	|_|	|_|	
UNIMPROVED								
g	Unprotected hand dug well	|_|	|_|	|_|_|	|_|	|_|	|_|	
h	Unprotected spring	|_|	|_|	|_|_|	|_|	|_|	|_|	

i	Surface water (river, dam, lake, pond, stream, canal, irrigation canal)	|_|	|_|	|_|_|	|_|	|_|	|_|	

Section 3: Schools Within a 30 Min. Walking Travel Time or 5KM Radius of Village

	3.1	3.2	3.3	3.4	3.5	3.6
School No.	School name	Type SEE CODES	*(IF 3.2=4)* Specify other.	In what year was this school established?	Levels taught SEE CODES	*(IF 3.5=5)* Specify other.
1		|_|		|_|_|_|_|	|_|_|_|_|	
2		|_|		|_|_|_|_|	|_|_|_|_|	
3		|_|		|_|_|_|_|	|_|_|_|_|	
4		|_|		|_|_|_|_|	|_|_|_|_|	
5		|_|		|_|_|_|_|	|_|_|_|_|	
6		|_|		|_|_|_|_|	|_|_|_|_|	
7		|_|		|_|_|_|_|	|_|_|_|_|	
8		|_|		|_|_|_|_|	|_|_|_|_|	
9		|_|		|_|_|_|_|	|_|_|_|_|	
10		|_|		|_|_|_|_|	|_|_|_|_|	

Codes for 3.2
1 Govt run
2 Non-profit, private
3 For profit, private
4 Other (specify)

Codes for 3.5
1 Nursery
2 KG
3 Primary
4 JSS
5 SSS
6 Other (specify)

THE INTERVIEW IS OVER. YOU WILL GATHER THE REST OF THE INFORMATION BY OBSERVING CONDITIONS IN THE COMMUNITY.

Section 4: Direct Observations

Community

	Question	1 Strongly disagree 2 Disagree 3 No opinion 4 Agree 5 Strongly Agree
4.1	The air in this village/township is generally polluted.	|_|
4.2	There are generally exposed garbage piles visible along the roads.	|_|
4.3	There are generally exposed cattle manure piles in this village/township.	|_|
4.4	There are blocked drains or obstructed river/gutter in this village/township.	|_|
4.5	There are standing pools of water (not including marshes, lakes) in this village/township.	|_|
4.6	The house yards in this village/township are generally not swept clean.	|_|
4.7	The grass growing in public squares in this village/township are generally not well cared for.	|_|
4.8	There are generally many flies visible (near the food vendors).	|_|
4.9	People are defecating in the bush	|_|

Village/Township/Township Welfare

	Question	0 No 1 Yes 7 Not applicable
4.10	In this village/township is there a NEIGHBOURHOOD Security Post?	|_|
4.11	In this village/township are there any signs/sign boards with public facility directions?	|_|
4.12	Does this village/township have a community centre?	|_|
4.13	In the market area in this village/township does one see shops or buildings that have been damaged or burned?	|_|
4.14	In the shopping district of this village/township does one see shops that have closed/gone out of business?	|_|
4.15	Are there any forest areas within this village/community?	|_|

Annex C:
Procedures for Identifying and Replacing Outliers in Water Data

Key variables of interest in this study include time spent by household members collecting water, volume of water collected from different water sources, and the cost of water among households. These variables are each derived from responses to several questions asked to respondents.

For example, the weekly time spent by a household collecting water from a water source was calculated using the following formula:

Where:

i=the water collector (we recorded information about up to two water collectors from each source);
t=the time spent travelling to and from a water source in one trip for collector i;
w=the time spent waiting in line at a water source in one trip for collector i;
f=the time spent filling containers at a water source for one trip for collector i; and,
n=the number of times per week that collector i collects water from the water source.

The formulas to calculate other key variables were of similar complexity. Missing data poses larger than normal threats to such variables because the absence of one input causes the entire equation to become incalculable. Even a small rate of missing values for each input can cause large rates of missing values for the calculated variable. (In the above equation, a missing rate of 2% for each of the inputs can yield a missing rate of more than 15% for the calculated variable.)

For these variables, outliers also pose larger than normal threats for parallel reasons – that is, if even one of the inputs is an outlier, it can cause the entire calculated variable to also become an outlier. The problem of outliers and missing data is compounded by the diversity of households in our sample. A water source was rarely used by a majority of households (i.e. many households "legally" skipped questions on many water sources). This means that there are small to moderate numbers of observations on many sources, implying that outliers and missing data have large effects on summary statistics.

Between outliers, missing data, and legal skips, sample sizes for key variables had the potential to quickly dwindle to unusable quantities. In order to counteract this, NORC recoded outliers using standard methods.
Outliers were identified for each water source separately. This is necessary because, for example, travelling 30 minutes to get to piped water inside a dwelling is likely to be an outlier, while travelling 30 minutes to collect water from a stream is unlikely to be an outlier. Outliers were recoded when there were 30 or more observations for a variable for a given water source; the 30 observation rule was imposed because of concerns about the reliability of summary statistics derived from smaller samples.[22] For each variable that was recoded, we first calculated the mean and standard deviation (SD) of that variable for each source. Then, we identified any values that were more than two standard deviations above the mean. These values were recoded so that they equaled two standard deviations above the mean. This is a standard method used to identify and impute outliers.

[22] Extreme values were, however, likely to be caught by the application of the universal rules described below.

The method was used on the following variables:

- Time spent travelling to water source
- Time spent waiting in line at water source
- Time spent filling containers at water source
- Number of containers filled per week
- Volume of container used to collect water (which is derived by container measurements or a printed volume on the container)
- Monthly price paid
- Price per liter per month

After recoding the above variables using the two-standard-deviations-above-the-mean rule, we applied several global recodes. These were applied to the variables across all water sources and regardless of the number of observations for a variable. These global recodes were used because, even after the above cleaning steps, there were several observations that defied logic. The global recodes used were:

- If the volume of water container used was less than five liters, we recoded it to a volume of 10 liters.
- If the respondent reported travel times or waiting times at a water source in excess of 120 minutes, the observation was recoded to equal 120.
- If the respondent reported that they spent 30 minutes or more fulling water containers, the observation was recoded to 30.
- If the respondent said they collected water from a given source 21 or more times per week, the observation was recoded to 21.
- If a respondent said they filled 15 or more containers per trip, the observation was recoded to 15.

In calculating the total volume of water consumed per household, we did not include rainwater collection because, even after taking the above cleaning steps, there were many large values for water consumed. This is because some households reported that large containers were used for rainwater collection, but that they collected water from this source many times per week (most likely using a small container to fetch water from the larger one that they collect it in). Therefore, it was impossible to estimate an accurate volume of water consumed from rainwater collection per week.