Evaluation Design Report

Ghana Water activity (MCC-13-BPA-0067-MCC-13-CL-0001)

Submitted to:	Submitted by:
Millennium Challenge Corporation (MCC) 875 15th Street NW Washington, DC 20005	The University of Notre Dame Notre Dame Initiative for Global Development (NDIGD) 926 Grace Hall Notre Dame, IN 46556 Phone: 574.631.2754

March 31, 2014

List of Acronyms

AT	As-Treated
ERR	Economic Rate of Return
FGD	Focus Group Discussion
GPS	Global Positioning System
IRB	Institutional Review Board
ITT	Intended to Treat
KII	Key Informant Interview
MCC	Millennium Challenge Corporation
MIDA	Millennium Development Authority
NDIGD	Notre Dame Initiative for Global Development
NORC	National Opinion Research Center
RCT	Randomized Controlled Trial
WHO	World Health Organization

TABLE OF CONTENTS

1. INTRODUCTION

The Millennium Challenge Corporation (MCC) is a U.S. Government Corporation created under Title VI of the Foreign Operations, Export Financing, and Related Programs Appropriations Act, 2004. The main mission of MCC is to provide assistance that will promote economic growth and elimination of extreme poverty in selected countries.

In August 2006, the Government of Ghana signed a 5-year $547 million Compact with MCC aimed at reducing poverty through economic growth led by agricultural transformation. This program was implemented by the Millennium Development Authority (MIDA), a government corporation established by the Parliament of Ghana to serve as the accountable agent for the implementation of the Compact.

MCC's aim is to rigorously evaluate the projects it funds to assess its investment impact and contribution to the development literature for knowledge dissemination. MCC is committed to using impact evaluation resources where they will provide the most useful lessons. Governments and organizations often design and fund projects where the link between the activity and poverty reduction is anecdotal. In these cases, rigorous impact evaluations can help establish, or refute, the links between costly investments and stated benefits. Also, MCC often funds similar projects in several countries and is interested in evaluating the effectiveness of these projects in different contexts. In some circumstances, governments may expand program(s) following the MCC investment using the result of the impact evaluation. In all of these scenarios, impact evaluation should provide lessons that will help focus limited funds where they can most effectively address development priorities.

NDIGD is collaborating with MCC on a rigorous impact evaluation of the Water activity in Ghana. The evaluation supports two objectives derived from MCC's core principles: accountability and learning. Accountability refers to MCC and MIDA's obligations to report on their activities and attributable outcomes, accept responsibility for them, and disclose these findings in a public and transparent manner. Learning refers to improving the understanding of the causal relationships between interventions and changes in poverty and incomes. The main purpose of the impact evaluation for the Water activity is to assess the impact the improved water systems have on beneficiary communities.

In line with MCC's two pronged principles, accountability and learning, answers to the following questions will be sought during the evaluation:

- Was the MCC investment implemented according to plan? What positive and negative factors affected implementation?
- Whether the investment in water systems improved health conditions for beneficiary communities?
- Did the MCC investment reach intended/unintended beneficiaries?
- Has the investment in the water sector decreased the distance traveled and the time households devote to acquiring water? If so, has the time freed through more efficient water collections shifted significantly to income producing activities?
- Has the investment led households to consume a greater quality of potable water for domestic purposes?

- Has the investment in the water systems resulted in increases in household consumption expenditures?
- Has the investment had different impacts for men and women?
- Has the investment decreased the price households pay for drinking water? Are households consuming a greater quantity of potable water?
- Are the benefits sustainable? Are the water points still in use?

A well designed data collection and evaluation can capture answers to these questions. NDIGD's prior experience in collecting data from communities using rigorous research methods is helpful in designing and executing this evaluation.

This document provides an overview of the design of the Water activity impact evaluation. Section 2 provides an overview of the Compact and the Water Activity. Section 3 reviews the literature on the impact of water projects on the well-being of the population. Section 4 explains the evaluation design proposed for this project. Section 5 describes data sources and outcome indicators and measures. Section 6 describes the analysis plan and section 7 describes the monitoring plan. The final section of the document provides information on ethics review, team roles, responsibilities, data quality and management.

2. Overview of the Compact and the intervention evaluated

The Ghana Compact was concentrated in 30 of Ghana's 170 districts and is comprised of 14 activities within three (3) focus areas. The Water component of the Rural Development Project's Community Services Activity was designed to provide improved water systems to 137 selected communities in the intervention area. Over the course of the compact, a total of 392 water points were constructed/reconstructed, including boreholes, small town water systems and pipe extensions.

Three types of projects were implemented in program communities,

- Boreholes fitted with hand pumps in small communities,
- Small town water systems comprised of lift system for water pump and distribution, and
- The Tamale water extension system in the north that extended the piped-water delivery system to surrounding communities.

2.1. Program Logic

The theory of change behind the water activity is that by improving water systems in districts participating in the Compact, the households' economic productivity and income will increase. This increase could reduce unproductive time spent caring for the sick, and/or collecting water. Since the quality of water will be improved by the project, the prevalence of illness, particularly diarrhea, will be reduced and the overall health status will improve. In addition, the water activity could help shift time formerly reserved for collecting water to income producing activities.

Rural households in developing countries need substantial amounts of time for water collection as each round trip from home to a water source can be lengthy and can require multiple trips (Meeks 2012). The MCC/MIDA Ghana performance indicator system documents the average distance for collecting water at 1,190 meters in the base period, and the project intervention aimed to reduce this distance by more than half (500 meters), By shortening the distance, time can be saved. The baseline report (p. 62) mentions, “all households’ members together are spending around four hours a day (240 minutes) collecting water.” That means water collection work is taking a considerable amount of time which could be utilized for other productive activities. There is little doubt that providing new water sources in the proximity will reduce the total amount of time people spend collecting water.

The link between safe drinking water and health outcomes is unquestionable. Water from unsafe sources is the cause of life threatening diseases in rural communities. Sources like unprotected springs, carts with small tank/drum, tanker trucks, and surface water (river, dam, lake, pond, stream, canal, or irrigation) are more likely to cause several types of diseases. The prevalence of these diseases is high in the locations where this project was implemented in Ghana; namely, diarrhea, Guinea worm[1] and Bilharzia. When people get sick due to water borne diseases, it costs time and money. When children get severely sick they need to be rushed to hospitals. That leads to economic costs for adults. WHO estimates that 88 percent of diarrheal disease instances are attributable to unsafe water supply, inadequate sanitation, and inadequate hygiene worldwide. The provision of improved water systems can help to reduce the risk of infection, which can help reduce the prevalence of waterborne diseases, saving time and money. Diarrheal diseases account for 12% of childhood deaths in Ghana, and are the third largest cause of death for children under the age of 5. These diseases are caused by the ingestion of water contaminated by fecal matter, and 20% of Ghana’s population does not use an improved water source (IPA 2008)[2].

3. LITERATURE REVIEW OF THE EVIDENCE

Literature suggests that improved water source can in general result in better health outcomes. Jalan and Ravallion (2003) demonstrate that the prevalence and duration of diarrhea among children in rural India were significantly lower for families with piped water. Utilizing a nationally representative household survey dataset, collected from 33,000 rural households by India’s National Council of Applied Economic Research in 1993-1994, the authors establish that the disease prevalence amongst those with piped water would be 21 percent higher without it (Jalan and Ravallion, 2003).

There have been some rigorous evaluation studies done that link improvement in water quality and health benefits. Evidence from randomized control studies have shown that improved water quality can reduce the prevalence of gastrointestinal diseases (Ahuja et al., 2010). Nonetheless, the evidence from randomized evaluation on increasing access to water without improving its quality does not confirm the link between water and disease reduction. Deveto et al. (2009) find that the provision of piped water to homes in Morocco did not have a significant effect on the incidence of disease. However, they did find that the provision of piped water saved time which was used for leisure and social activities. One reason given for not finding positive impact on health is that those sampled houses had accesses to relatively good quality water before the intervention so the new supply did not significantly improve the health (Deveto et al. 2009).

[1] However, recently Ghana government announced that Ghana eradicates Guinea Worm: http://abclocal.go.com/wpvi/story?section=news/health&id=8278438

[2] http://www.poverty-action.org/project/0125

There are some studies exist that look at health impacts of boreholes in rural areas. Curtis and Cairncross (2003) found that provision of drinking water from a borehole reduces diarrhea prevalence by 48 percent. Mbonye (2004) reports that water using from a borehole will be 2.2 times lower likelihood of diarrhea compared to drinking from unprotected and open sources like rivers and streams. There is strong evidence that large-scale investments in water and sanitation infrastructure can have massive impacts on child mortality (Zwane and Kremer, 2007). For example, Galiani et al. (2005) found that child mortality fell 5-7 percent where there was improved water system with improved service and expanded coverage in Argentina. Interestingly the effect was largest in the poorest areas, at around 24 percent.

Meeks (2011) in a study in Kyrgyz villages demonstrated that the introduction of water infrastructure benefited the villagers by reducing the time spent caring for sick children. This reduction in the time and intensity of home production allowed for greater time allocated toward leisure activities and labor market, especially working on the household farm. As a result, the average production of cash crops (specially, cereals such as wheat and barley) increased by 645 Kg per household per year in program areas (Meeks, 2001). However, many treatment communities in Ghana benefited from components of the compact's Agricultural Project, such as farmer training or post-harvest infrastructure or production credit. Hence, establishing a causal effect between water and agriculture production will not be possible since both were implemented during the same period.

When water is not available on the premises, women are more often responsible for water collection than men are. On average, an adult woman is the person usually carrying home the water in 63 percent of rural households in Sub-Saharan Africa (United Nations, 2010). Women in rural sub-Saharan Africa are the most burdened not only because they are usually the ones in charge of water collection but also because more time is needed in that region to bring the water home. The time needed to go to the source of drinking water, get water and return home is on average is 36 minutes in rural Africa (United Nations, 2010). While provision of new water sources, be it connected to the household or installed in proximity, results in less time spent collecting water, however it does not guarantee that those saved time will be eventually translated into time gain and involvement in productive activities (Meeks 2012). Devoto et al (2011), as reported above, found that the saved time was reallocated into the leisure activities in Morocco.

Infrastructure has been widely identified as one factor constraining women's economic opportunities. The provision of water could ease women burden and that might translate into higher off-farm activities in developing countries. However, in a study utilizing rural household and community data from national consumption surveys spanning Sub-Saharan Africa (Madagascar, Malawi, Rwanda and Uganda), South Asia (India, Nepal and Pakistan), North Africa (Morocco) and the Middle East (Yemen), Koolwal and van de Walle (2010) found no impact of water infrastructure on women participation in market based activities. However, the researchers found more support for the hypothesis that a reduction in the time to water has positive impacts on child schooling. For example, a one hour reduction in the time to water collection would increase girls' and boys' enrollment rates by about 8-9 percent in Yemen and by 18-19 percent in Pakistan. Interestingly, these impacts are found exclusively for non-African countries (Koolwal and van de Walle 2010).

The sustainability of the water project is of great concern, especially if there are not enough resources to maintain water points. Sustainability can be gauged in some part by the consumer's willingness to pay (Raje et al. 2002). The value a household places on a change in the supply of good quality water can be collected from the beneficiaries. Different formats can be adopted in a contingent valuation study. For instance, an open-ended format would directly ask questionnaire respondents what is the maximum amount they would be willing to pay for the improved water system (Genius et al. 2008). In a study, Khuc (2013) found that on average for all of the sampled households, the value of willingness-to-pay makes up a small percentage of household income, just 0.247% of total household income in Vietnam (Khuc 2013). The decision to pay for water depends on both internal factors: the level of education and awareness, as well as external factors: living conditions and existing water source. In another study, Genius et al (2008) found that the mean willingness to pay for improved water quality and quantity was estimated to be 17.67% over their water bill, which is equivalent to 10.64 € in Rethymno town, Greece. In India, nearly 50% of respondents were ready to pay partially more than their current bill amounts for improved quality of water (Raje et al. 2002).

Any information on willingness to pay for improved water system collected through this survey would provide valuable information on how much people value improved water projects which is important for the long-term sustainability of the water project and the increase in coverage of pipe water systems.

3.1. Evidence gaps that the current evaluation fills

The evaluation design established for this project will add to the literature on the impact of water improvement projects on health and economic activity. Specifically, this evaluation adds evidence on the links between the water improvement project and increase in quantity of water use, price of water and willingness to pay for water, time use and economic activity.

Researches that study links between time use, water improvement and economic activity have been doing so using a recall method. This method presents multiple measurement problems. First, in rural communities the concept of time as measured by hours and minutes might not be practical. Second, the process of recall activities might be biased depending on the respondent expectations, and how far back in time the activities can be recalled. People tend to under or over-estimate the time spent on certain activities. If it is a burdensome activity, like water fetching, people tend to exaggerate the time taken.

The design of this study attempts to improve the measurement of time by combining the recall questionnaire used in the baseline survey with following a person that usually collects water with a GPS instrument and a time use questionnaire. These two methods will provide not only a better measurement of time use, but will also help us understand the amount of bias introduced in recall methods. The introduction of GPS measurement has been piloted by the Notre Dame Initiative for Global Development in Burkina Faso in three villages. Among all the members of the household that collected water, one person was randomly selected to carry the GPS for as long as possible, usually 24 hours. The next day, the interviewers collected the GPS and applied the questionnaire as in the baseline. The GPS allowed us to see the distance to the fields, the amount of time spent in the fields and with the coordinates of each water point we were able to measure the time and distance spent collecting water. Some problems using the GPS were encountered during the pilot. Some people refused to use it either because they did not understand how the GPS worked, or in one case because it was thought to be a religious symbol. Among the comments received, the most interesting one was why we gave it to just one member of the family. It would be interesting to give it to a person that collects water and to a male, with preference to head of household. Gender differences could be estimated. However, this will be determined once we further analyze the data from the pilot.

Similarly, we also want to incorporate questions on willingness to pay so that we can estimate how much value the people place on the improved water system, and whether people who have received the water points value water more than people who have not received a water improvement point. This is especially of interest in areas where household can pay to be connected to the water distribution system. Anecdotal evidence indicates that people might be willing to be disconnected and continue with water fetching. Engel et al. (2007) report that 43% households that had access to improved water sources still using unsafe water in Volta Basin, Ghana. One of the reasons given for not using water from the improved source was price as higher the price lower will be the chance of getting water from improved sources.

4. Evaluation Design

The evaluation design was established by NORC, hence the end-line evaluation will continue with the same methodology developed by NORC to estimate the impact of the water improvement project. The baseline evaluation was designed to answer the following six hypotheses:

1. The incidence of diarrhea will decline, particularly in children 5 years old and younger.
2. Time devoted to acquiring water will fall significantly.
3. Time freed through more efficient water collections will be shifted significantly to income producing activities.
4. The price paid for drinking water will decline significantly, where water has been previously purchased.
5. Households will consume a greater quantity of water for domestic purposes.
6. Household consumption expenditures will increase.

This evaluation employs a pretest-posttest with pair matched comparison groups. In order to assess the impacts of the water points, the research design uses a difference-in-difference estimate of the program impacts. These estimates account for changes over time (before and after the intervention) in the treatment and control groups and for differences in those changes between the treatment and control groups. The measure of the program impact is thus a double-difference estimate. In this study, difference-in-difference estimates of the program impacts will be obtained for each of the main outcomes of interest as defined in the evaluation design (i.e., incidence of diarrhea amongst children 5 years and younger, time savings in acquiring water, water expenditures, water consumption, and income).

Note the differences-in-differences method of obtaining treatment impact estimates implicitly assumes the existence of four distinct groups of outcomes: 1) those that apply to the treatment group before the program, 2) those that apply to the control group before the program, 3) those that apply to the treatment group after the program, and 4) those that apply to the control group after the program. The "before" time refers to before completion of the water system improvements, and the "after" time is some time after completion of those improvements and a reasonable time for effects to occur. Note also that the "before" measures for the treatment and control groups are derived from the baseline survey that was already conducted.

MCC/MIDA supported rural communities to build/ expand/renovate water points. All of the water points were completed by January 2012. We assume that by 2014 the benefits of water activity might be visible and measurable in the communities. 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 completed and the end-line survey is conducted.

The quasi-experimental design used in this project will produce impact estimates of the water improvement projects. However, there are threats to the validity of these estimated impacts.

1. If there are factors that were not observable while determining a community to be in the treatment or control group, the impact might be upward biased. However, it depends on how it correlates with the dependent and independent variables.
2. If water improvements in control groups are performed because MIDA was "taking care of the selected communities", this positive externality might downward bias the impact of the water improvement project in the treatment communities. If an institution is looking to invest in water projects, they might exclude the communities where MIDA was working, because they might think there will be better return to investment in other communities.
3. Measurement error in any of the outcome or control variables might downward bias the estimate of the impact.
4. Attrition could be an issue since we are conducting a survey four years after implementation with the same households that were included in the baseline. Further, there is also a risk that the baseline characteristics may have changed over the two years between the data collection and completion of water project. We have a control group approach with difference in difference to address this issue.
5. Contamination between the treatment and control communities might have occurred, which is generally referred as compliance issue. This means a community that was supposed to receive a water point ultimately did not, and some communities that were designated as controls might not have received water points from MIDA. (See Annex 1 for the list of communities that have compliance issues). Further, if a control community received a new water project from other source, we handle this problem by calculating the Average Treatment Effect and the Average Treatment on the Treated. However, we are aware that this was not a randomized control trial.

To correct these potential threats to validity, we will introduce controls in the regression analysis that include characteristics of the community, and include tests for measurement error and unobservable characteristics. If needed, corrections using instrumental variables such as distance to water points will be used. However, there will be always remaining a risk of being different on some unobservable characteristics.

To limit attrition during the data collection, several measures can be adopted, for example: 1) sampled households will be informed that survey teams are coming and will be requested to answer enumerators' questions; 2) Information about the households will put in place including household contacts, location and GPS coordinates; and 3) we have an oversampling of households, estimating a 20 percent attrition rate for the second data collection.

We will collect data on additional replacement households in case the attrition rate for any community is larger than 20 percent. The analysis of this data will be used in two ways: a calculation on the impact of the project on the units for which we have two measures in time; and analysis at the community level by averaging indicators at the community level and performing the difference-in-difference estimator at the community level. Tests on the effect of attrition will be performed.

The contamination between treatment and control groups can be dealt with in multiple ways. The most common techniques used are: Average Impact on the Intended to Treat, and the Average Impact of the Treatment on the Treated. The first technique is considered the standard measure from the policy perspective. Outside of RCTs, policies sometimes will be adopted by one group and rejected by another. The second technique measures the impact of the Treatment on the Treated, and it might be a closer approximation to what the effect might have been without deviations from the protocol. However, this is only valid if the deviations to the protocol are not severe and biased toward certain unobserved characteristics. Usually, this analysis requires the use of an instrumental variable. It is common practice to use the initial random assignment as the instrument. In our case, however, this might not be possible since the treatment and control villages were not assigned randomly. However, we could use the fact that some villages could not receive the water points due to geological reasons. A more thorough explanation of these issues is provided in section 6.

In addition to the quantitative analysis, we will complement the evaluation with focus groups and key informants to inform the survey. We will include questions on spillover effects and unintended benefits to the key informants and community people.

4.1. Country-specific and international policy relevance of evaluation

One of the areas that this evaluation is interested in is assessing impacts of improved water system in the case of Guinea worm, as Guinea worm has in the past been a major problem in some of the communities in northern districts in Ghana. MCC/MIDA aims to reduce the incidence of Guinea worm by breaking the lifecycle of the worm that causes this disease. While Ghana has announced that zero cases were reported in 2011, the MCC/MIDA water project hopes to ensure that the disease is not reintroduced. Our evaluation will keep the questions on Guinea worm and collect information from the both treatment and control communities. However, our survey will have challenges to evaluate the Guinea Worm impacts. We understand that the detection of guinea worm requires extremely large sample size to give a reliable measure, this couple with a pressure from the government to agree that there is no guinea worm will likely to make the measurement even more difficult. Further, a measurement of guinea worm requires a different type of analysis that shouldn't be done through a household survey.

4.2. METHODOLOGY

Our method to evaluate the impact of MCC's investments in Ghana will integrate quantitative and qualitative methods to understand not only the direction and size of the impact but also the underlying mechanisms, processes, and channels through which the impact was generated. For this, in addition to survey, we will conduct a qualitative study such as focus group discussions (FGDs) and key informant interviews (KIIs) to gain a qualitative understanding of how the project has impacted the lives of people.

The evaluation design uses rigorous methods in data collection and analysis. Rigorous impact evaluations, which yield very strong evidence of program effects, involve answering a counterfactual question. The counterfactual answers the question, "What would have happened to program participants if they had not participated in the program?" This requires using a control group that is statistically identical to the treatment group receiving the intervention to act as a "stand-in" for units in the treatment group. This evaluation used a Propensity Score matching technique, a quasi-experimental design, in identifying and creating control communities. 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. As mention above, there are threats to the validity to determine a causal effect. Although the quasi-experimental design is the most rigorous in absence of an experimental design, designs that track participants over time (before and after, panel, and time series) cannot control for the effects of developmental changes that would have occurred without services, or for the effects of other events outside the program's influence.

4.3. POPULATION BEING STUDIED

The population that is being studied under this evaluation are the beneficiaries of the water activity. A total of 153,000 people living in 27,407 households in 30 districts of Ghana were affected by the water activity. This evaluation will be conducted with 600 households in treatment communities. We will have similar numbers from the control communities.

4.3.1. POWER CALCULATIONS, SAMPLE SIZE REQUIREMENTS

For statistical power analysis we assume that the main outcomes of interest have been defined at the household level and that assignment into treatment conditions has been performed at the community level (the clusters). Furthermore, we assume "blocked" assignment into treatment conditions, with the matched-pairs constituting the blocks. We are taking into account the main features that will determine the power to detect a treatment effect in the "pretest-posttest-with-matched-comparison-group quasi-experimental design," as described in the Baseline Report.

The statistical model used for the power ensuing power analyses is

$$y_{ijk} = \gamma_0 + \gamma_1 W_{jk} + \gamma_2 X_{jk} + u_{0k} + u_{1k} W_{jk} + r_{jk} + e_{ijk}$$

where y_{ijk} denotes the outcome for household $i = \{1, \ldots, n\}$ in community $j = \{1, \ldots, J\}$ in matched-pair $k = \{1, \ldots, K\}$; W_{jk} denotes a treatment-control contrast (or a treatment indicator) for community j in matched-pair k; X_{jk} is a (typically mean-centered) community-level covariate; γ_0 is the overall mean (intercept); γ_1 is the main treatment effect; γ_2 is the main covariate effect; u_{0k} and u_{1k} respectively denote matched-pairs and treatment-by-matched-pairs random effects; and r_{0jk} and e_{ijk} respectively denote community- and household-specific random effects. Our interest here is in determining the statistical power to reject the null hypothesis of no treatment effect ($H_0: \gamma_1 = 0$) in favor of the alternative hypothesis of a treatment effect ($H_a: \gamma_1 \neq 0$).[3] Note the model has been specified to account for a single control group and a single treatment group (i.e. 2 treatment conditions). Note also that for this power analysis the main outcome of interest has been implicitly defined as a continuous outcome. Finally, note the model contains four types of random effects: one to account for the variation between matched-pairs, one to account for the variation in the treatment effect across matched-pairs, one to account for the variation between communities and one to account for the variation within communities. There is however no effect (fixed or random) to account for the multiple measures being considered for each household.

The amount of statistical power is here determined by:

- The sample size at each level. More specifically, the number of households per community, denoted n; the number of communities per pair, denoted J (here $J = 2$); and the number of pairs (denoted K). We use n to denote the harmonic mean of households per community rather than the simple mean. This allows for a better approximation to the analytical model results. For the scenarios discussed below, we assume $n = 10$, $J = 2$ and $K = 50$. Varying n has a relatively marginal effect in the amount of power, but changes in either J or K have the potential to affect power in non-negligible ways. What really matters is JK, the total number of communities. In the scenarios discussed below we assume a total of $JK = 100$ communities. Since this is matched-pair design, we have one community in each pair in the treatment group and one in the control group (i.e. a balanced design). The design need not be balanced though. We can accommodate more communities in the control group relative to the treatment group.
- The (unconditional) intra-class correlation (ICC), denoted ρ. The ICC is the proportion of total outcome variance lying between the communities. In general, a larger ICC results in less statistical power as there is more variation between the communities to be explained. The ICC is dictated mainly by the context and the outcome measure being considered – it should not be selected based on any concerns related to the amount of statistical power available. Below we discuss $\rho = \{0.10, 0.30\}$.
- The proportion of outcome variance explained by the pairing, denoted B. Below we discuss $B = \{0.20, 0.40\}$. As with the ICC, the amount of outcome variance explained by the pairing is context and outcome specific.
- The proportion of outcome variance explained by a community-level covariate, here denoted R^2_{L2}. Below we discuss $R^2_{L2} = \{0.25, 0.50\}$. Although we consider a single covariate, the approach is quite general in that we assume a generic covariate that

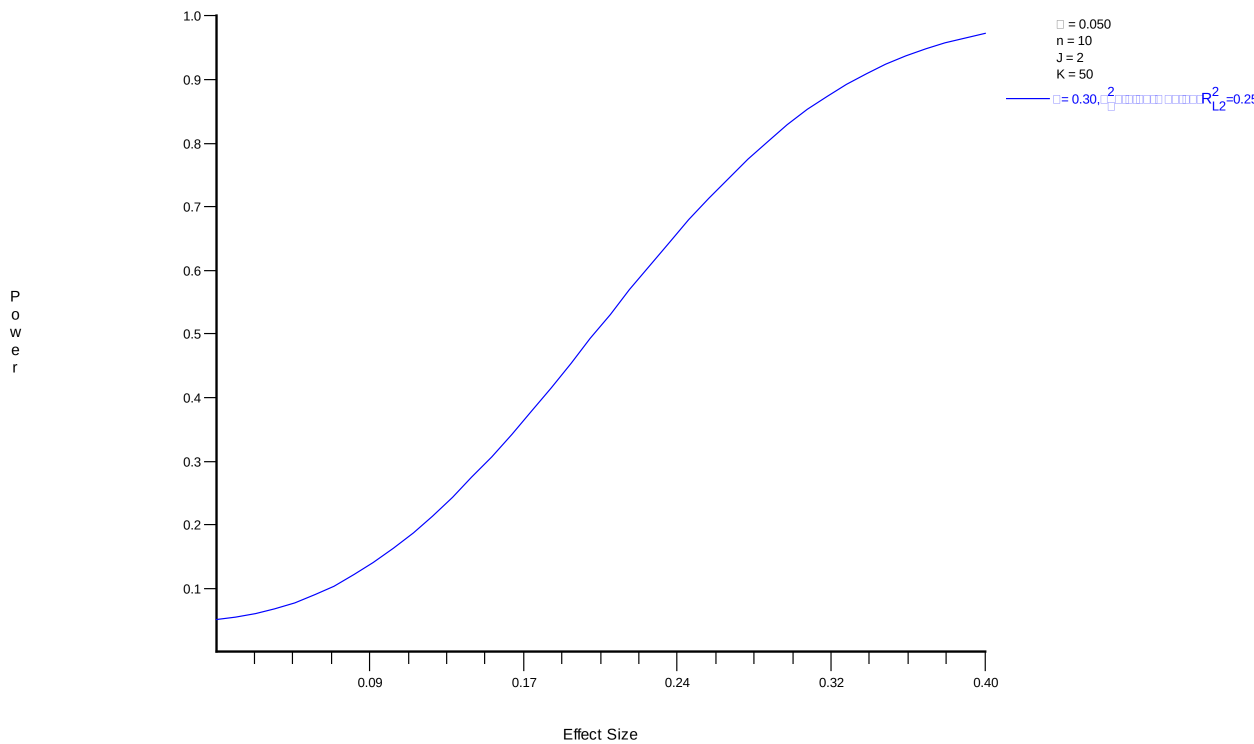
[3] A more formal specification of the model and the formulas used to calculate the statistical power can be found in the Optimal Design software documentation.

represents the strongest covariate available, with additional covariates adding only marginally to the proportion of variance explained whilst each using an additional degree of freedom.

- The effect size variability, denoted σ_{δ}^{2}. Below we discuss $\sigma_{\delta}^{2} = 0.01$. This is a measure of how much the treatment effect will vary between the matched-pairs.
- The type I error rate, here denoted α and hereafter assumed to be 0.05.

Figure 1 below shows the relationship between treatment effect size and statistical power assuming $\rho = 0.30$, $\sigma_{\delta}^{2} = 0.01$, $B = 0.40$ and $R_{L2}^{2} = 0.25$, $n = 10$, $J = 2$, $K = 50$ and $\alpha = 0.05$. As shown, the statistical power increases rapidly for treatment effect sizes between about 0.10 and 0.40. At 0.80 power the minimum detectable treatment effect size is about 0.29 and at 0.90 power the minimum detectable treatment effect size is about 0.34.

Figure 1: Treatment effect size vs. statistical power for a community pair-matched hierarchical quasi-experimental design with outcome at the household level and treatment at the community level and parameters as shown.



Suppose now a different ρ, B and R_{L2}^{2}. Figure 2 below once again shows the relationship between treatment effect size and statistical power for $\rho = \{0.10, 0.30\}$, $B = \{0.20, 0.40\}$ and $R_{L2}^{2} = \{0.25, 0.50\}$. All other parameters are as before. Note that the statistical power increases rapidly for treatment effect sizes of about 0.10 to about 0.40, regardless of the ICC, of the proportion of variance explained by the pairing, or of the community-level covariate. Note also that a larger B and R_{L2}^{2} result in additional statistical power for a given treatment effect size. Finally, note that a larger ICC results in smaller statistical power for a given treatment effect size. Treatment effect sizes at 0.80 and 0.90 power for each curve are shown in the table below. These effect sizes can be interpreted as the minimum detectable effect size (MDES) at the specified level of statistical power. Note the smallest MDES (0.21 at 0.80 power and 0.24 at 0.90 power) are obtained with the smallest ICC and the largest B and R_{L2}^2, whereas the largest MDES (0.29 at 0.80 power and 0.34 at 0.90 power) are obtained with the largest ICC and the smallest B and R_{L2}^2.

Figure 2: Treatment effect size vs. statistical power for a community pair-matched hierarchical quasi-experimental design with outcome at the household level and treatment at the community level and parameters as shown.

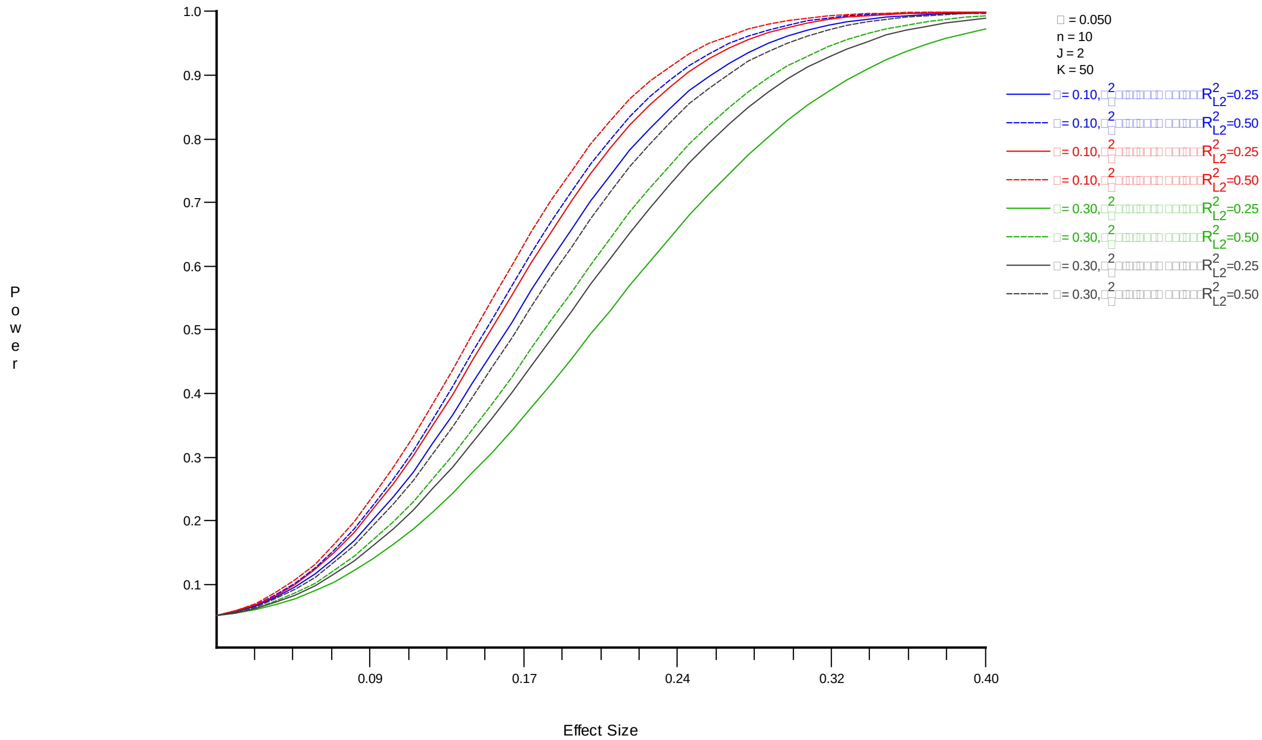


Table 1: Minimum detectable effect size at 0.80 and 0.90 power for a community pair-matched hierarchical quasi-experimental design with outcome at the household level and treatment at the community level $n = 10, J = 2, K = 50$ and other key values as shown.

			$R_{L2}^2 = 0.25$	$R_{L2}^2 = 0.50$
At 0.80 power	$\rho = 0.10$	$B = 0.20$	0.23	0.21
		$B = 0.40$	0.22	0.21
	$\rho = 0.30$	$B = 0.20$	0.29	0.26
		$B = 0.40$	0.27	0.24
At 0.90 power	$\rho = 0.10$	$B = 0.20$	0.26	0.25
		$B = 0.40$	0.25	0.24
	$\rho = 0.30$	$B = 0.20$	0.34	0.30
		$B = 0.40$	0.31	0.27

In summary, under the assumptions discussed, the MDES for this study will most likely fall in the range [0.21, 0.29] at 0.80 power and in the range [0.24, 0.34] at 0.90 power, and very likely toward the center of those ranges. The actual MDES will of course depend on how close the actual ICC and proportions of variance explained by the pairing and the covariate turn out. In general, larger ICCs or a smaller proportion of variance explained will result in larger MDESs.

4.3.2. SELECTION OF HOUSEHOLDS

Since this is a follow-up survey, we will track the same households that were included in the baseline. We got the complete list of households with contact information and the geographical position of the households from Baseline. Working with this information is very likely to at least help us locate the dwelling where the interviewed was done. If the family moved outside of the village, we will consider it a case of attrition. If the family can be found in the same village, we will try to find them.

4.4. TIMEFRAME

Preparation of detailed methodology:	Nov–Dec 2013
Reconnaissance visit:	Dec 1–10, 2013
Selection of survey firm:	Feb–Mar, 2014
Preparation of data collection manual	Mar–Apr, 2014
Training to the enumerators:	Oct–Nov, 2014
Qualitative data collection	June-July 2014
Survey:	Nov–Dec 2014
Data analysis:	Dec, 2014–Mar, 2015
Final report submission:	Apr–May 2015
Outreach activities:	Jun–Sept 2015

5. DATA SOURCES AND OUTCOME DEFINITIONS

NORC collected baseline data in 2010, so there is baseline data from program and control communities. Basically, the baseline data documents information household water consumption, members' health, time of collecting water, water price, consumption, etc. The surveys completed interviews with 1,200 households (12 households from each community) in 100 communities (50 treatment and 50 control communities).

5.1. QUANTITATIVE

This study will cover the same households that were used in collecting data in baseline phase for the impact evaluation. That means we will be conducting surveys in approximately 1,200 households in the treatment and control communities. As described in the Baseline Report, the impact evaluation has been designed as a "pretest-posttest-with-matched-comparison-group quasi-experiment". Furthermore, the Baseline Report Contractor used a community-level matched-pairs approach in determining the treatment and control communities. As we shall explain below, the proposed data collections methods and analytical model to be used (i.e. the statistical model used in the analytical stage) will appropriately and accurately reflect the design choices adopted, including the stratification and matching of communities.

Out of 137 program communities (where the water project was implemented), 50 communities were selected randomly during the baseline. These 50 communities served the treatment communities. Further, in the baseline 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.

During the baseline 1,200 households were selected to provide room for possible attrition with the aim to obtain a final sample size of 1,000. Both treatment and control communities will contribute equally in the sample size (600 each) as we will be conducting the survey in 12 randomly selected households.

Since the purpose of water project is to improve the quality and quantity of water supplied in the households, the unit of analysis will be households and we will collect data from adult members of the household that are present at the moment of the interview.

Questionnaire

Our questionnaire is based on the most appropriate and efficient approaches to answer the research questions. Of particular interest are effects on household health outcomes, time savings, and income levels. We will use the same questionnaire developed by NORC that was used in the baseline. Two types of questionnaires were used in the baseline—one for community and another for the households. We will make some modifications to the questionnaires as there are more research questions in the endline survey. For example, we are interested in knowing more about the sustainability of the water project, so some questions will be added to the community questionnaire:

Aspect of evaluation	**Questions (hypotheses) for impact evaluation**	**community questionnaire questions to be included**
Sustainability	o Are the water points still in use? o Are the benefits sustainable?	o Is there a water board in place in the small towns? o Is there a water and sanitation committee (WATSAN) in place in the villages with boreholes fitted with hand pumps? o Are the water points still in use? o Are people using the other sources? o Is there a trained caretaker to do repairs? o Is there an effort to mobilise funds for repairs and to ensure cost recovery? o Is the system managed in accordance with standard operation and maintenance guidelines?

Similarly, section 3 covers the household health in the last two weeks. The category “diarrhea” includes severe diseases, such as cholera, typhoid and dysentery – all of which have related “faecal–oral” transmission pathways. The respondents may not be able to properly name the disease and so the question has to capture all the symptoms. Based on that the type of disease could be evaluated, for example:

- Frequent evacuation, watery diarrhea might be an indicator of *e. coli* or cholera
- Frequent evacuation, bloody diarrhea might be and indicator of typhoid
- Highest frequent evacuation, with small stool at each visit might be dysentery

However, none of the symptoms are enough to determine a disease without a proper lab analysis.

Data collection

We will use mobile phones to collect data. Using mobile phone in data collection greatly increases the efficiency, minimizes the cost and reduces data errors. The survey questionnaire will be programmed in smart phones or tablets and the enumerators will collect the responses during the interviews using the phones. The data collected each day will be available for review and analysis the same day by uploading them to a secure server. All data collected is encrypted and can only be used by the researchers. In addition, because the computer measures start and end times of each survey, and can also collect GPS coordinates of the household, we can better supervise the work of the enumerators and reduce the possibility of data created by them and the productivity of each interviewer against the other enumerators.

We will guide the survey firm to ensure successful implementation of the evaluation work. We will first assess the firm's competencies in doing large scale surveys and their available human resources and their expertise. We will identify the areas where we can help them to develop their skills and arrange appropriate ways to do so. During the time of data collection the team members will be in the field and supervise day to day data collection work. Two members from Notre Dame and two additional members (local consultants) will monitor the work.

Instruments for the data collection will be thoroughly pre-tested and the methods used to obtain the data will be employed by only qualified and experienced enumerators in order to ensure a high quality data collection.

5. 2. QUALITATIVE

Qualitative data will be collected from focus group discussions and key informant interviews. The qualitative data will help to understand not only the direction and size of the impact but also the underlying mechanisms, processes, and channels through which the impact was generated.

Qualitative study is designed to elicit the personal and socially constructed realities of the community members, especially women who often carry the burden of collecting water. It will uncover the effects, meanings, behaviors and experiences of the water interventions in the lives of community members and households. While the results from the qualitative study may not be generalizable, it will provide a rich and deeper contextual understanding of the practical, nuanced and probably complex ways in which the new water infrastructure impacts the lives of men, women and children in the respective communities. Thus, the research process will be explorative and inductive, whereby patterns, models and trends are established based on the data collected on the identified themes as listed above. The research design will include gender analyses of the differential impacts of MIDA's water projects on the lives of men, women, girls and boys. The differential impacts of the water project among other socioeconomic subgroups (for example, age, marital and occupational) will be examined.

Data collection techniques will include in-depth interviews, focus group discussions (FGDs) and observations. Attention will be given to ethnography (language and local interpretations) in all these techniques to uncover the peculiar ethnic and cultural variations in how the water infrastructure affects people's lives.

We will collect qualitative data from 10 FGDs and 20-25 key informant interviews. The FGDs will be done in 10 communities. Each FGD will consist of six (6) to eight (8) members.

The FGD will be done based on following criteria:

1. All female
2. All males
3. Mixed (male and female)

We will organize 4 all-female FGDs, 2 all males FGDs, and 4 mixed FGDs. We will choose 10 communities purposefully (considering ecological zones, accessibility, prior water problem, type of existing water sources, etc). We make sure the participants of FGDs will be diverse in terms of geography (representing different locations of communities), occupation, age group, family size, users from different sources, etc. Some participants may be later participating in survey.

The FGDs will explore respondents' perceptions and experiences of the impact of the water projects on their own lives, and that of women, men, boys and girls in the community. Discussions will highlight the differential impacts that people of various demographic and socioeconomic backgrounds experience as a result of the water project.

In 10 communities, 2 key informants (1 female and 1 male) will be interviewed in-depth on the impact of the water projects in their communities on the lives of people. These individuals will also be in a position to discuss the challenges and issues involved in project implementation, factors affecting implementations, the management and maintenance of the infrastructure, conflicts and other administrative and management factors that could inhibit or enhance the utilization of the water infrastructure.

We will do interviews to collect information on whether the MCC investment also reached unintended beneficiaries. For example, in our recent trip to Ghana, we learned that some communities in Tamale were able to extend the pipe lines to other communities from the water point constructed by MIDA. We are particularly interested in learning how unintended communities might have benefited from the MCC investment. We will also inquire if any unplanned or illegal connections have been to the system that may impact the extension of pipelines.

Issues of location, utilization and sustainability of the water facilities among the beneficiary communities will be explored through direct observation. Photographs will also be taken under this component.

Basically, the observations will cover the following issues:

- Activities at water sites: who is fetching water? males? females? boys or girls? Are people happy or grumpy at the sites? Are there any complaints, expressions of satisfaction, or expressions of dissatisfaction?
- How is the water fetched at the sites? When is water fetched (peak and off peak times, what is the queuing situation? What is the convenience or inconvenience of fetching water)?
- What do people pay for water? Are there issues relating to charges, arguments over rates and other monetary issues?
- The location of boreholes, distance from households, accessibility etc.
- Any physically observable issues relating to water quality in terms of color and taste.

Interviews will be conducted in local languages, and the responses will be documented and also electronically recorded. Electronically recorded responses will be transcribed into English and word processed. A filing system will be developed to ensure that electronic copies of transcripts are properly stored for easy retrieval and we will make available of soft copies to MCC. Field supervisors will collate information and write reports on observations in the communities with note pads.

Data analysis will start from the day of community entry and through the data collection process.

The analysis of the transcripts will utilize interpretivism, thereby focusing on content and context, descriptions, language and narratives which reveal respondents viewpoints on the respective impacts of the water projects on their lives. Such emic perspectives show how they construct their reality with respect to the water project.

All typed transcripts will be read by the qualitative analyses team. A hierarchy of concepts and themes that are featured in the transcripts, supporting documents and the literature will be developed by the team and categorized. Systematically developed codes will be manually assigned to all concepts and categories. Soft copies of the transcripts will be manually coded. At this point, it may be possible to use either Microsoft Access or QDA Miner for sorting, filtering and generating results to establish patterns and identify factors that exert appreciable influence on outcomes. Electronically retrieved reports will be reexamined, referring to transcripts to establish context, interpretation and other factors that possibly underpin or influence the particular outcome. While the data analysis software will help establish patterns in responses with respect to key demographic and socioeconomic variables, content and interpretative analyses will reveal experiences, meanings, descriptions and contexts of phenomena. Narratives and snippets of interviews will also be used to illustrate findings from the study.

Gender analyses of qualitative data will be conducted taking into account gender relations, roles, norms, values practices and dynamics in the communities and how the MIDA water project has affected these constructs. We will compare all the results between treatments and control communities.

6. Analysis Plan

The impact evaluation design of the Water Supply Activity Program in Ghana was defined as a "pretest-posttest-with-matched-comparison-group" quasi-experimental design (NORC Baseline Report, p. 7). The following four groups were defined: a treatment group at two points in time, before and after, and a comparison group at the same two points in time. The before is "a time shortly before completion of the water system improvements", and the after is "sometime after completion of those improvements." The main outcomes of interest (e.g., time savings in acquiring water, water consumption and income) were defined at the household level with the main impact measure a double-difference estimate, also known as a differences-in-difference (DID) estimate, which is the difference in outcomes between the treatment and comparison groups of the differences between the before time and the after time.

One complication of the program being evaluated is that while the main outcomes of interest are defined at the household level, the treatment is implemented at the community level. That is, the main units of analysis (the households) are clustered within the units of treatment implementation (the communities). This clustering was addressed in the evaluation design by employing a two-stage sampling design. In the first-stage a sample of communities was selected. In the second-stage a sample of households was selected from within each sample community. Paraphrasing the baseline report: "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."

In the experimental design literature, units that are lost before the conclusion of the trial are often described as "drop outs", a situation that gives places to missing data. Also, units that do not stick to their assigned protocols (e.g., do not receive the treatment when they should have) are often referred to as non-compliers. While the design adopted in this program evaluation is not purely experimental, the community-level matching employed may be considered an approximation to an experimental design with group-level assignment into treatment and control conditions. As such, in this study any household in the baseline survey that for any reason cannot be included in the follow-up survey, should be considered missing data.

Missing data will have three important implications. First, the observed data will become unbalanced in the sense that the number of households may vary between communities, perhaps even substantially between matched treatment and control communities. Modern analytical methods can fairly easily accommodate this type of unbalance so it is something that can be addressed. However, observed data may also become unbalanced if entire treatment and control communities are lost in the follow-up survey, likely resulting in loss of information and perhaps some bias, as is discussed below.

Second, missing values can lead to some unavoidable loss in information, which can in turn reduce the efficiency of the study resulting then in less statistical power to detect the effects of interest. This would happen mainly because of entire treatment or control communities not being in the follow-up survey, but it can also happen because of a small number of households per community in the follow-up survey. Although household "replacement" (i.e., surveying different households in the same communities in which it is not possible to survey the same original households) will allow maintaining the initial levels of statistical power, one would have to be much more cautious about ascribing a causal interpretation to any difference found between the treatment and control groups. That is, household "replacement" would allow us to reach the intended levels of statistical power to detect an effect, but it could make it harder to call that effect the treatment effect. This ability to ascribe a "causal treatment effect" meaning to a difference, if any, between the treatment and control groups relates to the well-documented distinction between the intent-to-treat (ITT) analysis and the as-treated analysis (AT). A full ITT analysis is only possible if there is complete outcome data for all households included in the baseline survey.

A third implication is that missing values would likely introduce bias that can cause misleading inferences. The main problem that arises with missing data is that the distribution of the observed data may not be the same as the distribution of the complete data. Some missing values may be unrelated to the observed or unobserved responses, some may be related to the observed data, some related to the unobserved data, and some to both. The ITT analysis, which is widely accepted as an un-biased method for evaluation analysis, compares the study groups in terms of their treatment assignment, regardless of protocol deviations and participant compliance or withdrawal. Missing values lead to problems in identifying the ITT population, making the data analysis complex and challenging.

In summary, it seems that a full ITT analysis will not be feasible and the main challenge moving forward will be setting up an analysis that takes into account the bias, if any, that may be introduced by the missing data and by the replacement of households.

7. Monitoring Plan

We are aware that data collection work will be simultaneously carried out in different parts in Ghana. Mobilizing enumerators in different zones simultaneously requires a great deal of resources and careful planning. Ensuring the quality of the survey is another difficulty, as monitoring survey activities should not be compromised due to geographical challenges. We will ensure monitoring of survey activities through the provision of site supervisors, or by others means of communication if it is not possible to meet with all the enumerators personally.

NDIGD experts will identify data that are subject to erroneous or intentional alteration, manage a secure data room for restricted data designing, populating, and maintaining the social science research data, and will continually assess the data management needs for this project. State of the art technology will allow the researchers to determine and implement an appropriate analysis through data visualization. We will utilize encryption methods in our field work to prevent changes in the information, and we store our data in encrypted spaces in computers that can be accessed only by authorized personnel.

For quality control, the work of each interviewer will be closely supervised. Each day, whenever possible, experts from University of Notre Dame will check the data uploaded for any issues, especially for error, skipping and missing data. If anything suspicious is found, the NDIGD experts work with the interviewer to remedy the issue.

We will develop the appropriate statistical routines for internal consistency and validation checks. We will then review the data to be sure the checks were done correctly and will develop additional checks during early stages of data analysis.

8. ADMINISTRATIVE

8.1. SUMMARY OF INSTITUTIONAL REVIEW BOARD REQUIREMENTS (IN-COUNTRY, INTERNATIONAL)

We have already obtained an IRB approval from the University of Notre Dame for the survey work in Ghana. All evaluations conducted by NDIGD are subject to the University of Notre Dame's IRB. The IRB's primary role is to safeguard the rights and welfare of all human subjects who participate in research. In compliance with Federal law and institutional policy, all research projects involving human subjects must be reviewed and approved by the IRB. All social and behavioral research or biomedical projects conducted by the faculty, staff and students of the University are subject to the Policies and Procedures of the Institutional Review Board. All faculty and students engaged in such research submit requests for IRB approval prior to beginning their work. This requirement applies to all such research, regardless of the source of support. The University of Notre Dame IRB adheres to the principles established by the Belmont Report and federal policies, including the United States Department of Health and Human Services (DHHS) announcement in the Federal Register (Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection (May 12, 2004) and Office for the Research Protections' (OHRP) (Title 45 part 46 of the Code of Federal Regulations). Additional guidelines regarding the ethical conduct of research exist within academic disciplines, such as The American Psychological Association's Ethical Principles of Psychologists and Code of Conduct and the Code of Ethics of the American Anthropological Association. The IRB is composed of faculty volunteers affiliated with the variety of disciplines at Notre Dame.

In addition to the approval of the Notre Dame IRB, we will seek approval from an approved IRB Ghana review board. Specifically, we will apply to the Noguchi Memorial Institute for Medical Research Institutional Review Board (NMIMR-IRB) at the University of Ghana. The Noguchi Memorial Institute for Medical Research Institutional Review Board (NMIMR-IRB) was established in 2000 as part of a requirement for collaborative research with a US funded project (NIH). This establishment took effect in consultation with the College of Health Sciences and the University of Ghana. The NMIMR-IRB is an independent and competent body set up to review, evaluate and decide on the ethical merits of research protocols ensuring and guaranteeing the rights, dignity, safety and protection of all individuals and communities who participate in research activities. The IRB reviews both internal and external research protocols received from NMIMR, School of Nursing, School of Public Health, School of Allied Health, University of Ghana Medical Schools, private organizations, individual researchers and protocols outside the Ghana.

8.2. Data Access, Privacy and Documentation Plan

We will place great emphasis on data protection and privacy. The interviewers will input survey response into password protected instruments (cellphone/tablets). Each day, whenever possible, we will download the survey onto the computer, check the data for accuracy and then erase the data from cellphone/tablets at the end of the entire data collection. The computer that holds identifying data will be kept secure by the University of Notre Dame experts in password-protected files inside of encrypted spaces.

The GPS location and the name of the respondent for each household will be recorded for purposes of data collection and analysis. Each household will be assigned an identification number generated by the team. Then name and GPS coordinates will be stored separately the household survey dataset so that it will be difficult to trace any individuals.

Data gathered from the qualitative method will also kept in secure files. There will not be any identifiable information that links to the person who gave the information. We will not document the names of the person who provide quotes during the focus group discussions.

We plan to use STATA in the analysis so we will transfer the household survey data into a STATA file. We will also prepare write-ups for the data file that documents survey date, community, enumerators, and keeps all the information secured.

8.3. Evaluation Team roles and responsibilities

In order to best implement the proposed evaluation, NDIGD has brought together a team of highly experienced experts in designing and implementing impact evaluations in developing countries, with extensive experience in Sub Saharan Africa. The team is comprised of the following personnel:

Dr. Lila Kumar Khatiwada, Program Manager, has implemented evaluation surveys in Ghana, Nepal, Kenya, Mozambique, Namibia, and the United States. Currently, he is working as a Monitoring and Evaluation Specialist at NDIGD. As Program Manager, Dr. Khatiwada, will be responsible for managing and mobilizing the team, coordinating with the stakeholders in preparing detailed evaluation plan, communicating with MCC, selecting survey firm, training the survey firm, prepare key question for FGDs, overseeing the survey activities, and preparing and submitting the deliverables.

Dr. Juan Carlos Guzman, Sr. Analyst/Impact Evaluation**,** has worked in a variety of research institutions around the world, including the World Bank. Currently, he is working as a Monitoring and Evaluation Specialist with NDIGD. As a Sr. Analyst/ Impact Evaluation, Dr. Guzman will be responsible for assessing evaluation design, preparing the sample frame, assisting in trainings with the survey firm, supervising data collection, documenting and analyzing data and preparing reports.

Dr. Edwin Michael, Sr. Analyst/ Water, is a Professor of Biology at the University of Notre Dame. His research interest is in epidemiology, disease modeling, and in public health. As a Sr.

Analyst/ Water (health), Dr. Michael's contribution will be important from public health perspectives. He will be responsible for assessing evaluation design, reviewing survey questions, preparing questions related to health, analyzing data, interpreting results, and preparing reports.

Dr. Sampson Oduro-Kwarteng, Sr. Analyst/ Water**,** has 15 years of experience as a Civil and Water & Sanitation Engineer in Ghana. As a Sr. Analyst/ Water, Dr. Oduro-Kwarteng's contribution will be important from a water and sanitation perspective. He will be responsible for assessing evaluation design, reviewing survey questions, preparing questions related to water and sanitation, training enumerators, analyzing data, interpreting results, and preparing reports.

Dr. Deborah Atobrah, Faculty at the University of Ghana, is a qualitative researcher, gender expert, and a native of Ghana. Dr. Atobrah will assist in preparing qualitative data collection instruments, provide training to survey firm, supervising FGDs, and analyzing qualitative data, interpreting gender disaggregated data and preparing reports.

Dr. Joe Kaboski, is a Professor of Economics at the University of Notre Dame. Dr. Kaboski is a quantitative expert and assists the team in designing and executing survey and analyzing data. Of particular note, he will support the team in reviewing and calculating ERR.

Mr. Andres Martinez, Research in the Center for Social Research at the University of Notre Dame. Mr. Martinez is a statistician, and will be responsible for documenting and analyzing data and interpreting the results.

REFERENCES

Ahuja, A., Kremer, M. and Peterzon, P.2010. Providing Safe Water: Evidence from Randomized Evaluations. *Annual Review of Resource Economics*. 2:237-56. DOI: 10.1146/annurev.resource.012809.103919

Curtis V, Cairncross S. 2003. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *Lancet Inf Dis* 3: 275–281.

Deveto F, Duglo E, Dupas P, Parienté W, Pons V. (2012). Happiness on tap: the demand for and impact of piped water in urban Morocco. American Economic Journal-Economic Policy, 4, 68-99.

Engel,S., M. Iskandarani, and M. Useche. 2007. Demand and supply of improved water in Ghanaian Volta Basin. Intl. J. of River Basin Management , 5(1): 31-36.

Genius, M., E. Hatzaki, E. M. Kouromichelaki , and G. Kouvakis, S. Nikiforaki, and K. P. Tsagarakis. 2008. Evaluating Consumers' Willingness to Pay for Improved Potable Water Quality and Quantity. Water Resour Manage (2008) 22:1825–1834.

IPA. 2008. Source: http://www.poverty-action.org/project/0125

Jalan, J. and M. Ravallion. 2003. Does piped water reduce diarrhea for children in rural India? Journal of Econometrics 112: 153-173.

Khuc, Quy Van. 2013. Household's willingness-to-pay estimation For safe drinking water: a case study in Vietnam. Thesis submitted in partial fulfillment of the requirements For the Degree of Master of Science Colorado State University Fort Collins, Colorado.

Koolwal, Gayatri & van de Walle, Dominique.2010. "Access to water, women's work and child outcomes," Policy Research Working Paper Series 5302, The World Bank.

Kremer, M., Leino J., Miguel E, Zwane A (2009). Spring cleaning : rural water impacts, valuation, and institutions. The Quarterly Journal of Economics 126: 145-205. DOI 10.1093/qje/qjq010

Mbonye, A. K. 2004. Risk Factors for Diarrhoea and Upper Respiratory Tract Infections among Children in a Rural Area of Uganda. *Journal of Health, Population, and Nutrition, 22*(1), 52-58.

Meeks, R. 2012. The economic impact of water infrastructure. Harvard Kennedy School.

NORC. 2011. Impact Evaluation of Water Supply Activity: Baseline Report. MiDA Contract: 4101106-01.

Pattanayak et al. 2007. Informing the Water and Sanitation Sector Policy: Case Study of an Impact Evaluation Study of Water Supply, Sanitation and Hygiene Interventions in Rural Maharashtra, India. Working Paper 06_04. RTI International.

Raje, D.V., P.S. Dhobe, A.W. Deshpande. 2002. Consumer’s willingness to pay more for municipal supplied water: a case study. Ecological Economics 42: 391-400.

Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.

United Nations. 2010. The world’s women 2010: Trends and Statistics. Department of Economics and Social Affairs.

Zwane, P. and M. Kremer. (2007). “What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review.” World Bank Research Observer; 22(1): 1 - 24.