*Updated Evaluation Design

MCC Water and Sanitation Projects in El Salvador

Miguel Almanzar†, Maureen Cropper‡ and Raymond Guiteras§

University of Maryland- College Park

Consortium led by Social Impact

Revised: April 2014

* We would like to acknowledge Kristin Komives, Christine Poulos and Subhrendu Pattanayak who wrote the initial impact evaluation design proposal

† Miguel Almanzar is a Senior Research Analyst at the International Food Policy Research Institute (IFPRI) and a Research Associate at the University of Maryland. m.almanzar@cgiar.org

‡ Maureen Cropper is a Distinguished Professor of Economics at the University of Maryland and Senior Research Fellow at Resources for the Future. cropper@rff.org

§ Raymond Guiteras is an Assistant Professor at the University of Maryland. guiteras@econ.umd.edu

i. List of Acronyms

ANDA	: Administración Nacional de Acueductos y Alcantarillados National Water and Sewerage Administration
DID	: Difference in Difference
DIGESTYC	: Dirección General de Estadísticas y Censos General Directorate for Statistics and Census in El Salvador
EHPM	: Encuesta de Hogares de Propósitos Múltiples Multi-purpose household survey
ERR	: Economic Rate of Return
FOMILENIO	: Fondo del Milenio Millenium Fund
FUSADES	: Fundación Salvadoreña para el Desarrollo Económico y Social Salvadoran Foundation for Economic and Social Development
GOES	: Government of El Salvador
IHSN	: International Household Survey Network
ITT	: Intention to Treat
MCC	: Millennium Challenge Corporation
MDE	: Minimum Detectable Effect
MSPAS	: Ministry of Public Health
M&E	: Monitoring and Evaluation
PSM	: Propensity Score Matching
WSS	: Water and Sanitation Sub-Activity
WHO	: World Health Organization
UNICEF	: United Nations Children’s Fund

ii. Table of Contents

iii. List of Tables

iv. List of Figures

1. Introduction

In November of 2006, the Millennium Challenge Corporation (MCC) signed a five-year, $461 million Compact with the Government of El Salvador (GOES) to improve the lives of Salvadorans through strategic investments in education, public services, agricultural production, rural business development, and transportation infrastructure. The Government of El Salvador set up a management unit called FOMILENIO to implement the five-year Compact from September 2007 to September 2012. Social Impact (SI) has been contracted by MCC to conduct an impact evaluation of the Water and Sanitation Sub-Activity (WSS) under the Human Development Project of the Compact.

The goal of the WSS was to enhance access to water systems and to improve sanitation services to the poorest inhabitants of the Northern Zone of El Salvador. The component was designed to provide piped water or (in a few cases) public taps for households that previously did not have access to this level of service and latrines to all water project participants[5] who did not already have improved sanitation.

The Millennium Challenge Corporation expects the water and sanitation interventions to:

- Increase household income by at least 15%[6];
- Reduce morbidity from water-related illnesses; and,
- Reduce the time and cost spent on seeking or purchasing water.

Reductions in water-related disease and time spent collecting water are in turn expected to lead to reduced expenditures on health care and increased attendance at school and work. In the economic analysis prepared for the water and sanitation component of the Compact[7], three-quarters of the expected benefits are attributed to reductions in coping costs, specifically the time costs associated with collecting water and the monetary cost of relying on alternative water sources (such as vendors) and storage systems.

The objective of this document is to delineate the different aspects associated with developing a rigorous impact evaluation of the WSS. The benefits of the water and sanitation projects will be measured with a rigorous non-experimental design that incorporates matching, a panel survey, difference in difference estimation, and econometric analysis. The sample size for the panel survey is powered to measure changes in one primary indicator of household welfare (household expenditure). The evaluation will also examine changes, albeit not necessarily with the same degree of precision, in coping costs, productive time use, diarrheal disease, school attendance, and access to and use of water and sanitation infrastructure. To the extent possible, we examine the distribution of benefits and outcomes across gender and socio-economic groups. Data collection for the panel survey will be done

[5] In this document, 'participant' or 'beneficiary' is used to indicate any individual in a treated household or treated community. People with access to potable water and / or sanitation as result of the intervention of the MCC compact.

[6] This was revised from 10% in the latest M&E plan. See MCC (2012)

[7] Available at : http://www.mcc.gov/pages/countries/err/el-salvador-compact

by the General Directorate for Statistics and Census in El Salvador (Dirección General de Estadísticas y Censos - DIGESTYC).

This document updates the initial evaluation design report to account for the changes that occurred during the implementation of the WSS. The purpose is to gauge how these changes affected the original design and to determine what can be done to improve the design in response to these changes.

2. Overview of the Compact and the intervention evaluated

The El Salvador Compact entered into force in September 2007 and ended in September 2012. The Compact consisted of three Projects with the collective goal of stimulating economic growth and reduce poverty: Productive Development ($68 million); Human Development ($89 million) and Connectivity ($269 million). The Human Development Project consisted of two Activities, the Education and Training Activity (28.7 million) and the Community Development Activity (60.3 million). The Community Development Activity consisted of three Sub-Activities: Rural Electrification Sub-Activity (31 million) , Community Infrastructure Sub-Activity (11.7 million) and Water and Sanitation Sub-Activity (17.6 million).

The water and sanitation infrastructure projects were the main component of the WSS which consisted of the construction of potable water and sanitation systems, technical assistance for community capacity building to ensure system maintenance and sustainability, and community education related to appropriate health and sanitation practices. The goal of the WSS was to improve the lives of the poorest inhabitants of the Northern Zone of El Salvador through enhanced access to potable water systems, by improving quality, reliability and building new systems; and improving sanitation services for the households in these depressed areas.

2.1. Program Logic - Input, output, outcomes and ultimate impact[8]

2.1.1. Compact-level

The overall logic of the compact was to improve the lives of Salvadorans in the northern zone. As such, the compact entailed a myriad of infrastructure development coupled with technical assistance aimed at connecting northern El Salvador with the rest of the country, helping create opportunities for the region's residents through increase access to markets, thru the east-west highway; increase access to electricity, thru expansions of the electrical grid and distribution of solar panels; increase access to water and sanitation facilities to decrease disease burden in the region; among other interventions in education, agriculture and other productive activities.

[8] Following Bosch et al. (2000) we define outputs as the direct products of the program, such as the number and extension of the projects build, training sessions conducted; outcomes as changes in behaviors and knowledge because of the projects, such as measurable increases in water quality and quantity, access to sanitary services; and impacts as the gains experienced by project beneficiaries as a result of the project, for example changes in income and expenditures, changes in morbidity, etc.

The Northern Zone of El Salvador contains half of El Salvador's poorest municipalities and suffered more damage from the country's internal conflict during the 1980s than any other region. Economic and social indicators in the Northern Zone are currently worse than the national average: In 2007, 44.7 percent of households in the Northern Zone were poor, compared with the 34.6 percent national estimate; 17.2 percent of households in the region lived in extreme poverty in 2007 compared with 10.8 percent at the national level. Human capital development is also lower in this region than in any other. The average level of schooling in El Salvador was 5.9 years in 2007, while the average in the Northern Zone was only 4.3. The percentage of illiterate people in the Northern Zone was 18.3 percent in 2007 versus an 11.1 national average[9]. The goal of the Compact is to reduce rural poverty by increasing regional economic growth through a five-year program of strategic investments and technical assistance in various sectors.

2.1.2. Project-level

The Water and Sanitation Sub-Activity consisted of the construction of 45 potable water and sanitation systems, technical assistance for community capacity building to ensure system maintenance and sustainability, and community education related to appropriate health and sanitation practices.

Initially the water projects would involve providing water to households that did not previously have improved services. However, the project application, selection process[10], and feasibility studies generated a more diverse set of projects. Twenty-five of the projects installed water in communities that did not previously have improved water services. The remaining twenty projects extend an existing water system to additional households, improved an existing system, or both improved and extended an existing deficient system.

Table 1 shows the outputs of the projects and Table 2 the distribution of beneficiaries across departments and type of projects. The projects consisted of 272,151 meters of pipes that connect 26 wells and 19 water sources to an estimated 32,929 beneficiaries[11] with 7,624 new metered household tap connections providing access to new/improved water distribution systems. In addition, 15 community taps where constructed.

Forty-three of the projects include the construction/improvement of composting or improved-hole latrines construction. This part of the WSS built 1,702 composting latrines, 108 composting latrines and repaired 212 latrines in the households that were to be connected to the water system. Forty-four of the projects include the construction of grease-traps to dispose of gray waters. In total 7,142 new grease traps were constructed.

[9] Source: Dirección General de Estadística y Censos (2007)

[10] See Annex 2: Project Selection Criteria for the selection process

[11] Beneficiaries are defined as people that where provided a connection to the new/improved water and/or sanitation system. This is about 8,168 direct beneficiary households.

TABLE 1 WATER AND SANITATION OUTPUTS SUMMARY

Beneficiaries	32,929
Pumping Line (mts.)	57,022
Distribution Network (mts.)	211,541
Total Mts. pipes installed	272,151
Wells	26
Sources	19
Metered taps	7,416
Grease Traps	7,142
Number of water meters	7,624
Composting Latrines	1,702
Improved Hole Latrines	108
Improvement of Composting Latrines (existing)	212
Public Taps	15

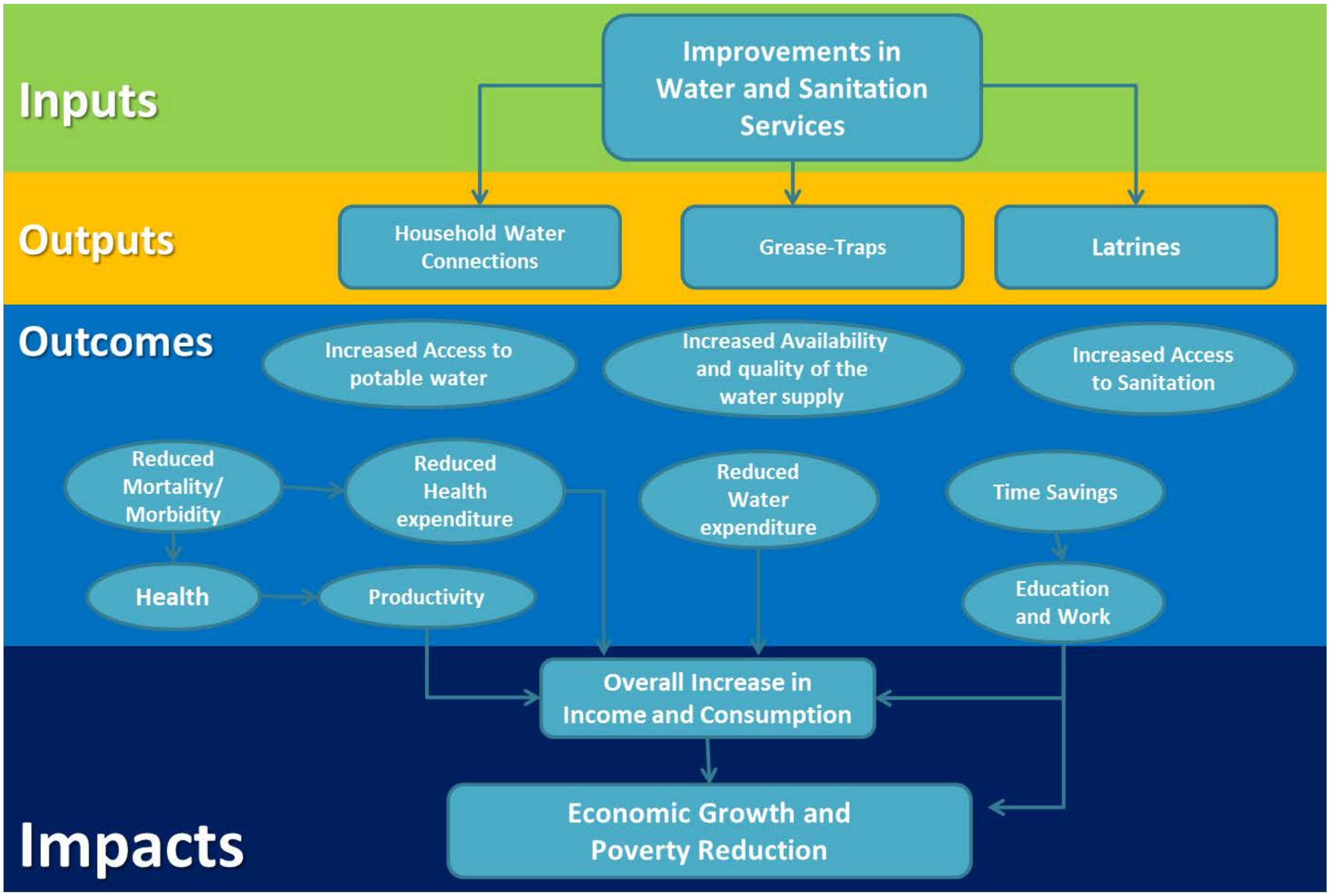
TABLE 2 NUMBER OF BENEFICIARY HOUSEHOLDS BY TYPE OF PROJECT AND DEPARTMENT

	Extended Systems	New Systems	Improved Systems	Total
CABAÑAS	230	135		365
CHALATENANGO	1,273	3,473		4,746
CUSCATLAN	574	2		576
LA UNION	230	1,018	45	1,293
MORAZAN	402	359		761
SAN SALVADOR		311		311
SANTA ANA	116			116
Total	2,825	5,298	45	8,168

Given the community focused nature of the WSS, an additional output of the WSS is the provision of training for the community staff to improve water management, both environmental and financial to ensure sustainability.

Figure 1 shows the impact pathways that relate inputs, to outputs, to outcomes and finally to impacts to achieve the overall objective of the compact. The principal outcomes of the WSS are the improved access to potable water and sanitation infrastructure, the improved availability and quality of the water supply. These outcomes are reflected in indicators such as: the proportion of households with piped water and sanitation facilities and the distance to the water source to measure access; cubic meters of water consumed and hours of service to measure availability; and, chlorination rates at the source to measure the quality of the service and of stored water to measure household storing behavior.

FIGURE 1 IMPACT PATHWAY OF WSS



The impacts on the well-being of the beneficiaries that the MCC expects the water and sanitation interventions to have are:

- Increase household income/consumption by at least 15%;
- Increased potable water consumption to 177 liters per person per day;
- Reduce morbidity from water-related illnesses, for example reducing diarrhea rates from 8.5% to 0%; and,
- Reduce the time and cost spent on seeking or purchasing water, for example reducing average time collecting water from 4.58 hours per week per household to 0 hours per week per household, and the cost of water from 1.68 $ per cubic meter to 0.43$ per cubic meter.

Other potential impacts of the WSS included improvements in education, measured as attendance and enrollment of children originating from decreased coping cost of carrying water an doing laundry outside the home and the decreased incidence of water-borne diseases; impact heterogeneity across gender and socio-economic status will be explored.

The project will offer benefits in terms of reduced incidence of disease caused by the currently sub-standard levels of water and sanitation service in the region. For example the baseline survey shows that around 45 percent of the households in the treatment segments use untreated water sources like wells and springs to obtain water and over 40 percent have unimproved sanitary services, namely nonexistent, communal and hole latrines. Beneficiaries gain time savings by removing the need to fetch water and cost savings by removing the need to buy water from more expensive sources. These benefits are expected to have positive effects on household incomes in the region. In addition, reduced mortality and morbidity entail specific benefits like reduced expenditures on healthcare, and potential labor productivity gains from increased attendance to school and work.

2.2. Link to ERR and Beneficiary Analysis

The Economic Rate of Return (ERR) measures the effectiveness of a program by contrasting the discounted flows of costs and benefits of a specific intervention. The costs are comprised of any initial investment and any required maintenance expenditures throughout the course of the program. The benefits are determined by the gains of the population affected by the project.

To calculate the costs and benefits, the initial ERR of the water and sanitation used data from the EHPM (Encuesta de Hogares de Propósitos Múltiples) 2004, the survey FUSADES - BASIS – 2003, water pricing information from ANDA, and statistics from the Ministry of Public Health (MSPAS) on water-borne diseases incidence. In this economic analysis, three-quarters of the expected benefits are attributed to reductions in “coping costs”, specifically the time costs associated with collecting water and the monetary cost of relying on alternative water sources (such as vendors) and storage systems. The initial ERR was 13.8% which was later revised down to 5.7% in 2010, and to 1.4% at compact close (as of August 2012)[12]. As noted the benefits are primarily time and cost savings from not having to fetch water/buy truck water.

The results from this study will serve to update the figures used in the ERR; first by updating the figures with the baseline data[13], and then by allowing a more flexible estimation of the costs and benefits using the follow-up surveys. The data for the follow-up serve to adjust the benefits estimates for the years of the survey and we can use the variation over the 3-year period to better predict the benefits and cost in the years after the projects.

3. Literature review of the evidence

The importance of providing improved water and sanitation as a way to promote development has long been recognized; from the effects on child mortality, to school attendance and work productivity gains, water and sanitation can improve the well-being of people throughout their life span [WHO/UNICEF (2005)]. Recognizing this potential, one of the millennium goals is to halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation. In

[12] Using a different methodology (adding half of the consumer surplus to account for the benefits of additional consumption) the ERR at compact close was 3.6%

[13] Closeout ERR uses the baseline survey data

the developing world, the lack of access to safe water and improved sanitation service has clear implications for the health status of the poor and also their economic life. Lack of access to water and sanitation not only exposes people to infectious waterborne disease, that decrease the probability of survival at young and old ages, it also imposes a burden in their economic life, by increases in time out of productive activities thorough illness, time fetching water and storage and treatment costs.

The health impacts of water and sanitation programs has been studied frequently [see Fewtrell et al. (2005); Pattanayak et al. (2008); Pattanayak et al. (2010); Newman et al. (2002); and Galiani, Gertler, and Schargrodsky (2005) Galiani, Gonzalez-Rozada, and Schargrodsky (2009); Devoto et al. (2012); Jalan and Ravallion (2003); Gamper-Rabindran, Khan, and Timmins (2010)]. Very few studies have measured other important outcomes, such as changes in the household's costs of collecting, storing and treating water, or the income losses due to water-borne and water-washed illnesses [Pattanayak et al. (2008)]. Furthermore, there are few rigorous impact evaluations that have measured education, gender, and poverty reduction impacts of water and sanitation interventions.

To give a brief overview of the rigorous impact evaluations in the water and sanitation sector, we follow Bosch et al. (2000) that categorized water and sanitation impacts on program participants into four groups:

1) health improvement;
2) education;
3) gender and social inclusion; and
4) income and consumption.

On the health side, impact evaluations have focused on child mortality. Specifically, given that diarrheal disease is the second leading cause of death in children under five years old and a leading cause of malnutrition [WHO (2013)], the health impact of water and sanitation have tended to focus on this measure. Newman et al. (2002) evaluate the investments in small water and sanitation projects in Bolivia, and find that community level training was needed to have effects on water quality, they also find effects of infant mortality; bringing forward the importance of "hardware" interventions being couple with "software" interventions to achieve goals. Galiani, Gertler, and Schargrodsky (2005) find that child mortality in Argentina fell 8 percent due to access to increased access to water (through privatization), with the poorest benefiting the most. Kremer et al. (2011) find positive effects in child health from a randomized experiment in Kenya that protected water sources, while Devoto et al. (2012) find effects on child health from an intervention that provided tap connections to an urban sample in Morocco. These two studies bring an important issue in the effects of water supply infrastructure: that it is important to distinguish between increases in *quantity* and *quality* of water. Kremer et al. (2011) evaluate the impact of an increase in the water *quality* available to treatment group by protecting the water sources, in the other hand Devoto et al. (2012) evaluate the impact of increasing the *quantity* of water available to households in an urban area, no changes in quality are expected in this setting since these households where obtaining water from this network before the intervention. Similarly, Devoto et al. (2012) do not find education impacts; however, this is probably due to children no longer being involved in water fetching activities. In rural Pakistan, Rauniyar, Orbeta, and Sugiyarto (2011) find that water and sanitation projects improved access to the water supply and improved attendance among girls; they find no effects on the labor supply or water-borne diseases.

Other studies explore the link between water and child health. Mangyo (2008) uses panel data and an instrumental variable[14] methodology finds that access to in-yard water source in China had positive impact on child health for children of educated mothers. Jalan and Ravallion (2003) find that piped water can lower the prevalence and duration of diarrhea among children under five in rural India using propensity score matching methods. Gamper-Rabindran, Khan, and Timmins (2010) use panel data to estimate a quantile treatment effect for the provision of piped water on infant mortality in Brazil. They find that piped water benefits are larger for areas with higher infant mortality. On the other side of the spectrum of the effects of health outcome, Klasen et al. (2011) evaluate the impact of increased access to piped water supply in Yemen and find that it worsens health outcomes when water rationing is frequent, likely due to pollution in the network. They find that connections to piped sewers can lead to health improvements but that these benefits are not cleared when compared to water supply through water vendors. Fan and Mahal (2011) find non-robust positive effects of water and sanitation on dysentery and significant reductions of diarrhea for children under five due to hygienic practices (hand washing).

An important issue in the literature is the complementarities of water and sanitation projects. For example, Esrey (1996) finds that improved water quality can improve child health if sanitation is also provided. Some rigorous evaluations like Pattanayak et al. (2009) and Pattanayak et al. (2008) using a similar methodology as the one proposed here, finds that a community demand driven water and sanitation intervention in India, had positive effects on the level of access to piped water and sanitary services, but no discernible health or education impacts.

On the effects of water and sanitation on income and consumption, gender and social inclusion the evidence is limited. Some studies find limited effects on these outcomes [Chase (2002), Lokshin and Yemtsov (2005), Kremer et al. (2011), Pattanayak et al. (2008), Pattanayak et al. (2010), Devoto et al. (2012)]. Impacts on consumption and income are achieved through changes in coping costs, both time and expenses, so that these impacts can be measured through indicators of expenditure on water and sanitation services, coping costs more generally as in Pattanayak et al. (2005), and total expenditure. Finally, impacts on gender and social inclusion refer to the extent that minorities, the poor or other vulnerable populations benefit from the water and sanitation interventions. The effects might be larger for some of these populations, because non-linear treatment effects, for example if women disproportionately participate in fetching water they would have larger benefits from a project that provides tap-water. On the other hand, if the contribution is large the poorer households might not be able to afford the connection and thus not benefit from the project even if targeted; in the same way aged households, are less able to provide a labor contribution and this vulnerable population can be excluded because of physical and economic constraints.

[14] Lamichhane and Mangyo (2011) warrant against the IV methodology due to endogenous project placement.

There are few studies that quantify the impact of water access on productivity in either agriculture or in the labor market. The mechanisms through which these effects can exist are evident: (1) decreases in time fetching water provides time that can be allocated to productive activities and (2) the decrease in water-borne illness provide for a healthier population that can be more productive. However, these effects have only been recently explored in the literature and to date no discernible effects are found, for example in Devoto et al. (2012) no changes in the time allocation to productive activities are found, and Koolwal and van de Walle (2013) does not find that access to water comes with greater off-farm work for women.

3.1. Evidence gaps that current evaluation fills

In this study, we will examine the effect of the Compact's water and sanitation sub-activity on the costs of collecting, storing and treating water, as well as the income losses due to water-borne and water-washed illnesses, which will constitute a significant contribution to the literature. This impact evaluation uses a rigorous quasi-experimental methodology to shed light on welfare and equity implications of water and sanitation infrastructure interventions.

We will not neglect the important finding in the public health literature that it is water quantity, and not necessarily quality that is of greater importance, by detangling the effects of the quality of water at the point of consumption in contrast to at the point of delivery [Esrey et al. (1991), Fewtrell et al. (2005)]. Economic studies have failed to pick up the distinction between quality at the source and quality at the point of consumption, even though these may differ dramatically. In addition, the variety of projects that were constructed could allow us to shed some light on the different impacts that can be expected from new water and sanitation system versus the improvement of the existent ones.

4. Evaluation Design

4.1. Evaluation type

The benefits of the water and sanitation sub activity will be measured using a rigorous quasi-experimental impact evaluation methodology. An impact evaluation is a study that measures the changes in outcomes that measure aspects of wellbeing which can be attributed to a specific intervention. Impact evaluations require a credible and rigorously defined counterfactual, which estimates what would have happened to the beneficiaries absent the project. Estimated impacts, when contrasted with total related costs, provide an assessment of the intervention's cost-effectiveness.

4.2. Evaluation questions

We divided the primary evaluation questions in different categories: welfare indicators, coping cost in cash and time, health, education, reliability and quality of service, spillover effects. In addition, we allow for differential impacts in gender and social groups for these main outcomes.

Household welfare

- Do water and sanitation infrastructure investments increase household expenditure or income? What factors might explain the impact (or lack of impact) in this area?

- What are the consequences of water and sanitation investments for expenditure patterns?

MCC and FOMILENIO expect the water and sanitation program to increase incomes by 15% among project participants. In this study, household expenditure will be used as a proxy for income and the major indicator of household welfare. We will also measure income, but in the interest of keeping the survey of reasonable length, we will focus on aspects of income that we believe would be most directly related to water and sanitation improvements and the time savings that such improvements could generate.

Because there is little information in the existing water and sanitation literature about the impact of interventions on income and because of the inherent difficulties in accurately measuring income and expenditure, this study will also attempt to study the mechanisms through which one might expect water and sanitation investments to generate income improvements. Water and sanitation investments could generate an increase in income through several channels:

- Working individuals could dedicate **more productive hour to income-earning activities** because, for example, they or their dependents are not ill as often or because they no longer have to spend as much time collecting water.
- Individuals who did not previously participate in income earning activities could **enter the work force** because they have more time available or because water and sanitation investments present new opportunities for them (e.g. the possibility of starting a water-related business in the home; new businesses open in the village).
- Current productive activities in the households (e.g. agriculture or small business activities) could **become more productive and/or profitable** with a more reliable and less expensive water source.

To explore these channels, we will gather information about time use in productive activities and water collection, productive time lost due to illness or care-giving and productive use of water at the household level.

In addition to (or in lieu of) increasing income and expenditures, water and sanitation investments could generate changes in patterns of expenditure. If households spend less on medical care due to reduced illness, or spend less on the purchase, storage, and treatment of water, then resources are liberated for other uses. If these resources are redirected to activities that contribute to socio-economic development (school fees, transportation, preventative health care, nutritious foods), then they could magnify the development effects of water interventions. To explore whether water and sanitation interventions produce changes in patterns of expenditure, we will examine changes not just in total household expenditure[15], but also in major categories of expenditure, such as food, housing and other.

Coping costs and cash expenditure on water

- Do water and sanitation interventions reduce coping costs? What factors might explain the impact (or lack of impact) in this area?

[15] Excluding water and sanitation cash coping costs.

- Do they reduce cash expenditures on water and on sanitation services? What factors might explain the impact (or lack of impact) in this area?

Coping costs are the expenditures that households make in order to collect, store, and treat water. Closer, more reliable, and better quality water is generally expected to reduce these costs. We will measure expenditures on building and maintaining alternative water sources, delivery systems, and storage containers. We will also measure time spent collecting water and washing clothes at a source outside of the home as well as cash expenditures on water services, water vendors, and bottled water. Additionally, we will examine cash expenditures on sanitation services (e.g. use of public latrines, expenditures on latrine emptying and maintenance).

One can expect the quantity of water used by households to increase with reliable piped service (e.g., see Strand and Walker (2005), Nauges and Strand (2007)). However, the availability of an improved water system has an ambiguous impact on the volume of water that a household consumes. A water system that places meters in households where no meters previously existed could in principle decrease water consumption if the water is priced at a higher level, while still having the level of consumption being optimal[16]. In addition, under block pricing, as shown in Olmstead, Hanemann, and Stavins (2007), households do not face an increasing marginal price for the water they consume. In this case, it might be optimal for the household to increase their water consumption to just below the quantity where the block price increases; essentially creating 'bunching' at the kinks of the price schedule and making the price a choice variable. The endogeneity of price under a block pricing schedule makes it difficult to estimate the water demand curve, Olmstead, Hanemann, and Stavins (2007); to this we can add the difficulty brought about by multiple source use (and the demand-choice implied). Instead of trying to estimate the water demand to evaluate the impact of the projects, we focus instead on estimating the impact of projects on water consumption patterns of treatment households.

Even if total expenditures on water do not change, the time and/or cash expenditure on water per unit of water used could decrease. Measuring the quantity of water used by households is very difficult. We will use questions about the size and number of containers of water collected, the storage units inside the home, and the time the tap is running to try to estimate changes in the quantity of water used by households before and after the interventions.

The survey is designed to measure water consumption across all the sources that a household has available. These include public taps, storage systems, hoses and others. In an effort to account for differences in behavior between households with a meter and does without, we will conduct the analysis using the sub-sample of households with metered tap connections. This will only include the water consumed from the tap even when they use other sources; albeit with lowered power to detect any impact.

Health

[16] As long as the shadow price of water without access to tap water is above the prices that one will pay with the water meter, consumption should increase.

- Do water and sanitation interventions reduce incidence of diarrheal illness?
- What factors (hygiene behavior, source and household-level water quality, household source choice) might explain the impact (or lack of impact) in this area?

While water and sanitation improvements are related to reductions in a variety of water-borne, water-washed and water-related diseases, most rigorous impact evaluations in the sector have focused on diarrheal disease and acute respiratory illness. This study will focus primarily on diarrheal illness, but also collect information about cases of respiratory illness, stomach ailments, conjunctivitis, dengue, and fevers.

The primary health indicator is a period prevalence of diarrhea, particularly in children under 60 months of age. Based on the predominant practice in the literature, the judgment of health experts and epidemiologists, and previous impact evaluations of water and sanitation projects measuring the health impacts [Pattanayak et al. (2008); Klasen et al. (2011); Klasen et al. (2012)], we define diarrhea as three or more episodes of liquid bowel movements within a period of 24 hours (with or without blood and/or mucus) during the two weeks prior to the survey.

We note that the sample is not designed to measure the health impacts caused by the interventions. The primary objective of the research design is to measure changes in expenditure and coping costs in a rigorous way. However, this study will rely on self-reported morbidity to measure health status. While self-reported disease measures are vulnerable to measurement error due to the fact that respondents may fail to recall episodes or misdiagnose disease, these are the predominate health indicators in water and sanitation impact evaluations and they are widely used by epidemiologists and health experts. It is important, however, to ensure comparability of measures taken in different years by conducting the panel household surveys in the same season each year. It is also important to control for other events that increase diarrhea, including weather events, natural disasters, and economic crises. In El Salvador, the start date of the rains is an important indicator of when diarrheal rates will begin to increase. We will, therefore, collect information on the start dates of the rains in each survey year and ensure that matched control and treatment villages are interviewed during the same week, and that all villages are re-interviewed as close as possible to same week of the year in follow-up surveys.

To address health in the evaluation, we combine two sources of information to estimate health impacts: information from the household survey and health facility data.

The health effects that are associated with water and sanitation interventions are produced through a variety of changes in household behavior and in living environments. To examine this, we will explore three issues that could contribute to the presence or absence of health improvements using the *household survey*:

- **Source choice and use of multiple sources**. For all households in the study, we will monitor self-reported use of community water sources to determine whether the introduction of an improved piped system leads to abandonment of other, unimproved sources.

- **Hygiene knowledge and behavior, water treatment, and use of sanitation facilities**. For all households in the study, we will monitor understanding of good hygiene behavior. We will also examine whether sanitation facilities are being used and hand washing practiced. To study water treatment behavior, we will monitor self- reported water treatment practices (boiling, filtering, and use of PURIAGUA [a chlorine additive]) as well as test drinking water samples for residual chlorine levels in all interviewed households.
- **Household and source water quality**. To assess the extent of bacterial contamination of water in households and at sources, the study will include laboratory tests of drinking water and source water in a sample of households and villages[17].

The *health facility data*, will detail the number of cases of water-borne and other diseases in the area of influence. These data will be used to detect impacts in the incidence of disease across the health facilities in the northern zone.

Education

- Do water and sanitation interventions increase school ***enrollment*** among children aged 7 to 12? And children age 6 to 18? What factors might explain the impact (or lack of impact) in this area?
- Do water and sanitation interventions increase school ***attendance*** among children aged 7 to 12? And children age 6 to 18? What factors might explain the impact (or lack of impact) in this area?

In the literature on water and sanitation it is commonly suggested that in situations where children - particularly female children - are responsible for water collection, a reduction in the time spent collecting water could yield higher school enrollment and attendance. However, few[18] rigorous studies have examined whether these expected benefits of system improvements materialize. To examine this question, the survey will ask about school enrollment and school attendance over the two weeks prior to the survey, and the reasons for non-attendance.

Service, use, and sustainability

- Were the water and sanitation projects implemented according to plan?
- Are the results from the activity expected to be sustained over time?
- Did the MCC investment reach intended/unintended beneficiaries?

While impacts – the fundamental change experienced by participants – are the focus of this study, we will also measure some intermediate indicators that provide information about whether a project has been successfully implemented, what new options the project has opened for beneficiary and non-beneficiary households, and whether service levels are maintained over the period of this study.

We will measure changes in:

---

[17] See "Annex 5: Water quality testing" for the details on the water quality sampling and specific tests.

[18] For example, Devoto et al. (2012) find no impacts on educational outcomes like, time spent doing homework, school completion and the degree of absenteeism

- Access (e.g., the average distance from participants' homes to a water source and to sanitation facility or location, proportion of households with piped water and improved sanitation facilities),
- Quality and reliability of services (e.g., the number of hours of service, frequency of breakdowns, risk perception, chlorination rates, presence of micro-organism in the water at source and point of use),
- Use (e.g. household use of water and sanitation services)

These are important outcomes to gauge the success of the projects. We will examine the improvement in direct outcomes. Namely, to examine the changes in access to water and sanitation services in treatment segments in comparison to control segments.

We will measure access to improved water and sanitation services in various ways. First, by measuring the change in the percentage of households that have a tap connection in their household (on premises); and/or an improved sanitation facility, such as a composting latrine or an improved-hole latrine. Second, we define a water score and a sanitation score that measure the quality/type of the best service a household has access to and changes in the water sources used are reflected in the scores. Households that have access to potable/tap connections post intervention will have improved scores even if they do not abandon their use of unimproved water sources. Lastly, improvements in access to water services are also reflected as decreases in time to specific sources, such as public, neighbors' and private taps.

To measure quality we will use microbiological tests of a sub-sample, to measure the presence of fecal coliforms and E.coli on the sources an at the point of consumption. In addition, we will test for chlorine levels in the full sample. The level of chlorine in drinking water is indicative of the appropriate water treatment at the source and at the household.

Gender and social exclusion

- Do the effect on health, education and access of water and sanitation interventions differ by gender or by expenditure levels (initial conditions)?
- What factors (hygiene behavior, source and household-level water quality, household source choice) might explain the impact (or lack of impact) in a specific subpopulation?

How the benefits of water and sanitation interventions are distributed across categories of participants is a question that has not received serious study through rigorous impact analyses. In this study, we will collect information on the gender and age of household members, as well as on the relative socio-economic status of households. This information will be used to analyze, to the extent feasible with the sample available, differences in selected impacts and outcomes across gender and socio-economic groups. In the case of gender, we will pay particular attention to differential impacts in areas of time use and school attendance. For socio-economic groups, we will look at differences in access improvements, use of clean water, as well as changes in coping costs and health, education and major welfare indicators.

4.2.1. Country-specific and international policy relevance of evaluation

Improving livelihoods in poor countries through the development of infrastructure for drinking water supply, sanitation and waste-water management has long been recognized in the development international community. As mentioned before, rigorous impact evaluations are few in the water sector, with increasing attention being place recently. Rigorous impact evaluations allow us to explore the effectiveness of these investments, promoting accountability and helping to improve results in future projects of this nature. Given the costs in terms of time and money, only in a few cases are impact evaluations incorporated in the project's design. Given these constraints it is important to invest the impact evaluation budget to study cases where one is hopeful that the situation will lead us to a better understanding of the mechanisms through which these investments affect the livelihood of the poor.

The MCC WSS in El Salvador is one of these opportunities. First, the rural water sector in El Salvador has lagged behind, with only 48% of people having access to water sources on the premises and 53% having access to improved sanitation services in 2011 [WHO; UNICEF (2013)]. These proportions are not abysmal, thus providing a setting that is comparable to other countries in the region, which can serve to better understand how these investments can benefit participants in countries where there is some access to water and sanitation in rural areas.

The study contributes to the rigorous evidence on the impact of rural water and sanitation supply which is available in academic and policy discussions. Thus allowing better grounding of these discussions and improving the effectiveness of these investments, by showing what things work and which do not. Based on the findings, we will be able to draw some policy conclusions to provide input for designing future interventions related to investments in providing piped access to water in regions affected by water scarcity, the importance of the reliability of water supply and the households behaviors that need to changed and/or promoted to maximize the impacts of water sector investments.

4.2.2. Definition of key outcomes linked to program logic

The impact evaluation uses three waves of household surveys and community surveys; and one wave of retrospective health post survey. Table 12 shows a list of the primary indicators that will be analyzed. In addition, Table 13 shows the definitions and targets of the primary indicators in the monitoring and evaluation (M&E) plan [MCC (2012)]. These indicators will be estimated using the panel household surveys.

4.3.Methodology

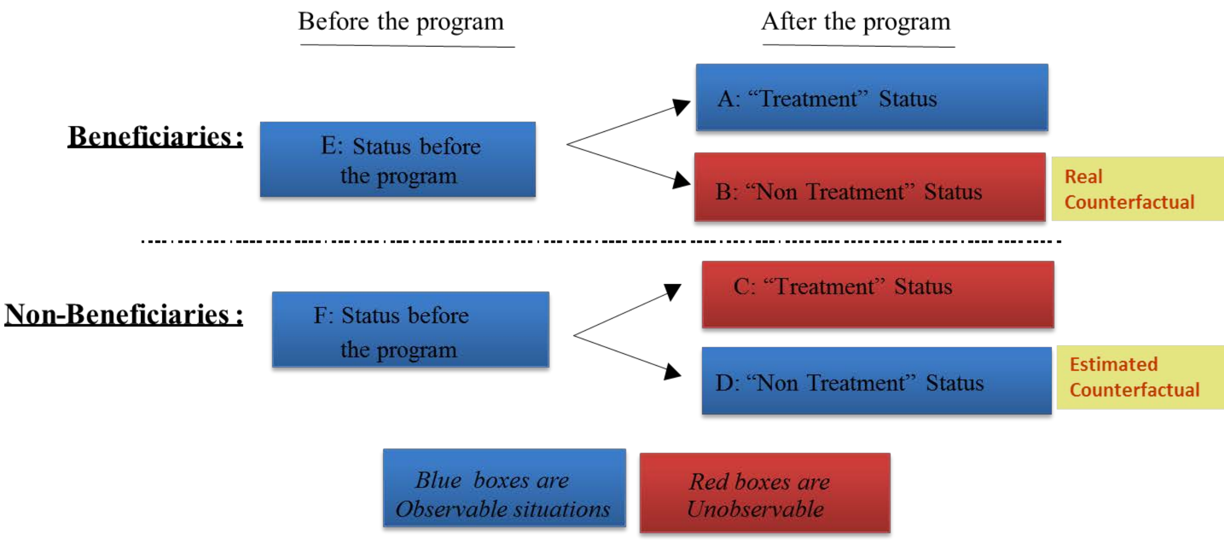
The key to measuring the impacts caused by the water and sanitation interventions is to compare conditions with the interventions to conditions that would have prevailed without them. The counterfactual state is not naturally observable – we can never know what change would have occurred in program participants (the treatment group) if the program were not implemented. As it was not possible to apply randomization in the selection of water and sanitation projects in this case, the benefits of the water and sanitation projects will be measured with a rigorous quasi-experimental design that incorporates matching, pre- and post-implementation data collection, difference-in-difference estimation, and econometric analysis to estimate the counterfactual and address selection and other biases. This requires selecting a comparison group—households that are observationally similar to beneficiary households but do not participate in the program—and observing both sets of households before and after the program is implemented.

Matching represents a credible non-experimental option for identifying comparison groups. We use propensity score matching (PSM) [Rubin (1974); Rubin (1979); Rosenbaum and Rubin (1983); Dehejia and Wahba (1994); Heckman, Ichimura, and Todd (1997); Heckman, Ichimura, and Todd (1998); Heckman et al. (1998)] using data from the 2007 census to match the treatment communities to comparable communities before program implementation. PSM identifies comparison communities that have a similar probability of receiving the treatment and are similar to the treatment communities in terms of observable characteristics. Accordingly, they provide measures of indicators in communities that are similar except for the treatment; thus addressing selection on observables.

By collecting data before and after program implementation, we can apply a 'difference-in-difference' (DID) estimator [Heckman, Ichimura, and Todd (1998)]. This estimator measures the treatment effect as the difference between the changes in indicators before and after the program among treatment recipients, on the one hand, and the changes in indicators before and after the program among comparison units, on the other. DID estimation helps control for residual confounding due to imperfect matches and selection bias from time-invariant unobservable factors which differ between treatment and comparison communities and which may have an influence on the impact variables of interest.

The basic principle of the DID estimator is the comparison between situations with the program and without the program, also known as "treatment effect". This approach is illustrated in Figure 2. In the impact evaluation, we would like to compare changes in those who were treated vis-à-vis what would have happened had these treatment not been provided (i.e. boxes A and B). This is opposed to a mere comparison of the situation after the program and before the program (i.e. comparing A to E or the difference between participants and non-participants (A to D)). Unfortunately, it is not possible to observe state B.

FIGURE 2 POSSIBLE SITUATIONS FOR TREATED AND CONTROL HOUSEHOLDS



To address this problem of unobservable situations, we identify a comparison group (D) that is as similar as possible to the treatment group using propensity score matching, so that observations of Dare a close approximation of the counterfactual B.

To implement the DID estimator, we have planned for a panel survey in which the same households are interviewed in 2011[19], 2012 and 2013. Implementing a baseline and two follow-up surveys has a number of advantages over an evaluation design utilizing only two data points. The longer one waits to conduct the follow-up survey, the greater the risk that the measured impacts will be due (at least in part) to changes that are unrelated to the program under study and that household will forget key details about project implementation. There is also the risk that some control communities will receive water and sanitation services from a different funding agency or program. Also, multiple data points allow us to adjust the sample frame to unexpected changes in projects and provide us with an interim and full-term view of project impacts; full-term in the sense that by the endline survey all projects were finalized and connected household were benefit from the increased access to potable water. With a follow-up survey conducted both one and two years after the baseline we both minimize risk of contamination of comparisons and gain a view of full-term impacts[20].

We will employ regression-adjusted DID estimation Ravallion (2008) in order to control for individual and household level covariates, with adjustments for intra-cluster correlation due to design effects. Information on the covariates is collected in the survey instrument.

4.3.1. Treatment Assignment

[19] There was an initial baseline survey done in 2009 based on feasibility study information. However, some of the project locations changed and implementation was delayed, so an "updated" baseline was conducted in 2011. We are evaluating the possible comparisons of the 2009 baseline with the follow-up surveys.

[20] In addition, we expect a wide range in the timing that the projects will be finalized. This opens the opportunity to exploit this variation to estimate a continuous treatment effect ("months with improved water").

Projects

There are numerous different ways to define the "treatment" that is applied in the WSS. The simplest approach is simply to say that the sum total of all activities undertaken in any project constitutes the treatment, thus ignoring the differences between the interventions in each community. If treatment is defined in this way, then the treatment effect that we measure with the evaluation will be an average effect across all types of interventions.

At the other extreme, we could consider defining four different treatments, as shown in Table 3 below. The implication of defining multiple treatments, however, is that it requires a research design that measures the program impact for each type of treatment, which in this case would mean close to doubling (in the case of two treatments) or quadrupling (in the case of four treatments) the sample size required for the survey. Aside from sample size considerations, there are some practical difficulties involved in precisely defining the treatments, as some of the beneficiaries of the expansion and improvement projects where not connected before , making it for them essentially a new water project.

TABLE 3 POSSIBLE WAYS TO DEFINE TREATMENT

	New Water Projects	Expansion and Improvement Projects
Sanitation	A	B
No Sanitation	C	D

Of the potential differences in treatment effects that we could analyze in this study, we consider that the most policy relevant and least studied in the existing literature is the potential differential effects between access to a new piped potable water system and the access to improved quality water systems. Communities in Latin American developing countries now find themselves in need of what could be called a third stage of water supply improvements, where quality and reliability take a center stage. They have passed the first stage of installing public point sources and a second stage of installing individual piped connections, namely increasing access to potable water. The third stage of improvements needs to address system failures and population growth; households are not receiving the quality and reliability of service they demand, or some new households are excluded from connecting to the existing systems.

MCC and FOMILENIO have decided to include this third generation type project in the WSS, but the rate of return calculations for the project are based on the expected benefits of new projects. For MCC and FOMILENIO, it is thus an interesting question whether the two types of water projects have similar impacts on income, school attendance, health, and coping costs. The comparison could also be interesting for the government of El Salvador as it plans for future water supply investments. We therefore see distinct advantages in building a test of two different treatments into this evaluation:

• Treatment A = “New” water projects (with or without sanitation)

• Treatment B = Projects that “improve, rehabilitate, or expand” existing systems (with or without sanitation)

To do independent tests of these two treatments, however, would require a larger sample size. Given concerns about costs of the evaluation and uncertainty about the final project designs and locations, we have decided to focus on a design that will estimate the benefits of the full FOMILENIO WSS as a single treatment (without differentiating between project types). Nonetheless, within the limits of the final sample size, we will explore possibilities for evaluating differential impacts of the WSS interventions.

Households

“Treated” households – or households that have received the treatment -- are defined as those households who had the potential to benefit from these water and sanitation projects. This means that they live within the service area of the new or improved water (and/or sanitation) system. Households that live within the service area, but do not connect to the service or see no change in their water or sanitation service after the project (e.g. water reliability, quality, hours of service from existing connection did not improve) are still considered treated households. This approach to defining treated households has important implications for the impact evaluation: projects that do not inspire households to connect or to agree to build latrines will have lower impacts than projects that achieve full coverage. One of the aspects of the project that is being evaluated is the “take up” and use of the services that is offered.

In defining treated households, we are also assuming that at least the short and medium-term benefits of community water and sanitation improvements are only enjoyed by households living with the community. If a household moves out of the community, its members cease to benefit from the latrine or water connection they received. Households that have moved out of the intervention area between the baseline and follow-up surveys will therefore be dropped from the panel study.

Communities

Treated “communities” are difficult to define in this case. The concept of community in rural El Salvador is associated with caseríos, which are residential areas that have a collective identity. Caseríos are not formal administrative units and do not exist in all areas. Each caserío is part of a “canton” [21], and a group of cantons forms a municipality. Most of the WSS projects are intended to serve a single caserío, but some projects will serve multiple caseríos. Thus, “project areas” are not equivalent to either caseríos, cantons, or municipalities in this water and sanitation intervention, and the water systems are not necessarily based in one “community”. [22]

In addition to the conceptual problems involved in defining community, we face a practical constraint in defining treatment communities for sampling purposes: there is no information available about

[21] With some caseríos falling in more than one canton

[22] This has implications for the management of these systems, as many systems will require collaboration of people from different “communities.”

households at the casería level or at the project area level. The lowest level at which census information is gathered is the census segment[23]. Each canton has one or more census segments[24]. In some cases, census segments are larger than project areas, and in some cases they are smaller. According to the information we were given before selecting the sample, census segments generally include about 100-125 households[25].

In this study, we ultimately chose to define treatment "community" at the level of the census segment. There are a number of advantages to defining community at the census segment level rather than at the project level or canton level:

- While "project area" would in many ways be the ideal choice of treatment community, we do not have the equivalent unit in control communities[26].
- A canton is usually much larger than a project area and will include many more non-beneficiary households than a census segment will.
- 2007 census data is available at the census segment level, so we had significant information available for matching treatment and control communities at this level.
- The General Directorate for Statistics and Census (DIGESTYC) is used to working with these census segments, and thus we are able to make use of many of their established procedures for sample selection and survey implementation.

4.3.2. Identifying the Treatment Segments

To define our sample universe we first identify the census segments that are expected to contain WSS projects. To do this, FOMILENIO staff and staff of the Census division searched the census data base to find the census segments where the caseríos, barrios, lotificaciones, or cantons included in the project descriptions[27] were located.

This proved to be quite difficult for a number of reasons. First, the information on caseríos is not systematically recorded in the census database. Rather the name of the caserío appears in the address field, only if a household reported the caserío when asked for the address. Some caseríos simply did not appear in the census data base. Others appeared but were associated only with a very small number of households. The list of identified treatment segments is thus likely to be missing some parts of project areas.

[23] This is the primary sampling unit for sample surveys in El Salvador. Its boundaries do not correspond to any political boundaries in most cases.

[24] In theory, each census segment is located in a single canton, but in practice there is debate about canton boundaries. We found therefore that a number of census segments span several cantons, at least according to the residents of the area.

[25] The rural census segments had about 120 houses, but the recent census showed that up to 20% of those houses can be unoccupied. For the purposes of the evaluation design, we assumed 100 households per census segment.

[26] We did make an effort to identify, through FISDL, WSS projects that have been formulated, but not submitted for FOMILENIO funding. This could have served as control "projects" for the study. However, there were not enough formulated but un-submitted projects to have this be the non-treated pool for the matching.

[27] See "Annex 3: List and description of projects"

Second, in some cases, the information from the census database did not match the information in the project descriptions, for example, because according to the census database the caserío where the project was to take place was located in a different canton than was listed in the project description. FISDL, a GOES agency assisting in the implementation of the project, was asked to check these cases. In most cases, their reply was that the project description was correct. In a few cases, they corrected the project description.

In the end, a census segment was considered a "treatment segment" if any households in the segment listed the project community in their address. Nine projects had to be removed from the treatment sample frame because we were unable to identify any census segments associated with them.

In the original design for the 2009 baseline, the plan was to draw a random sample of treatment segments. However, as it turned out, there were not many more treatment segments than the number required for this design; there were 100 treatment segments in the sample frame and the power calculations indicated we needed at least 88. We chose instead to use clear criteria to reduce the set of possible treatment segments to the 88 that we needed to comply with the power calculations at the time. Segments were eliminated if:

- The project associated with the treatment segment was very small (had fewer than 100 participants)
- The segment had fewer than 55 identified households (which from DIGESTYC's past experience means that it would be hard to select a sample of 18 occupied houses)
- The segment was also included in either the connectivity or northern zone multi-purpose household survey samples (which would make selection of households for the water survey difficult) and elimination of the segment would not eliminate the entire project area from the survey sample.

Elimination on these criteria left a total of 90 segments. To reach 88, we also eliminated one treatment segment from each of the two largest project areas (the treatment segment from each project that contained the smallest number of households). This process of systematic elimination did not significantly alter the distribution of segments across project types or across departments.

For the 2011 baseline survey, the list of segments that were expected to receive treatment was updated and the propensity scores were recalculated. Accordingly, the segments do not necessarily overlap with the 2009 baseline sample, except for treatment segments where projects were expected to be implemented in 2009 and are were still expected in 2011. In reality, the resampling of segments due to projects being dropped did not change the sample frame considerably, as expected; many of the segments with dropped projects were selected as comparisons by the propensity score procedure. The 2011 sample reduced the sample to 65 treatment segments and 65 comparison segments, as discussed in the sample design section.

4.3.3. Selection of comparison segments

Comparison segments were selected prior to the intervention using propensity score matching (PSM) [Rosenbaum and Rubin (1983); Dehejia and Wahba (1994); Heckman, Ichimura, and Todd (1997); Heckman, Ichimura, and Todd (1998) Heckman et al. (1998)]. The PSM procedure allows us to select a control unit j that is very similar to a treatment unit i in all observable aspects except for treatment status, thus providing a proper counterfactual of the situation of i without the treatment. A "propensity score" is the probability that a given unit participates in the program given certain observable characteristics Z. Thus, the propensity score is given by the probability that a unit is treated conditional on having observed the set of characteristics Z, that is $P(T = 1|Z)$.

$$P(T = 1|Z) = \Phi(Z\hat{\alpha})$$

where the function $\Phi(.)$ is the cumulative normal distribution (probit) or the logistic distribution (logit) and $\hat{\alpha}$ are the estimated coefficients.

We implement propensity score matching (PSM) using data from the 2007 census to match the treatment communities (segments) to comparable communities (segments) before program implementation.[28] Once we have estimated the propensity score function, we estimate the propensity score $P(T = 1|Z)$ for all treatment segments and all possible comparison segments. We assign an appropriate comparison segment to each treatment segment in the sample using nearest neighbor matching without replacement. Specifically, we find a segment *j* among the segments with T=0 for each segment *i* with T=1, so that the pair *(i, j)* satisfies

$$\min_{j\in\{T=0\}} \|P_i - P_j\|$$

According to the PSM method, the comparison segment *j* is selected such that its probability of participation in the program is as similar as possible to beneficiary i's participation probability. Intuitively, PSM creates the observational analogue of a randomized control group in which comparison units and beneficiary units have the same probability of participation.

Although propensity score matching can ensure that treatment segments are compared to non-recipients who are similar in terms of probability of treatment, there still may be both observable and unobservable differences between treatment and comparison segments that may bias the results. To eliminate these biases we use a difference-in-difference (DID) approach to control for "selection on unobservables" Heckman, Ichimura, and Todd (1998).

The comparison group was determined by matching the chosen treatment census segments with similar census segments located within the municipalities that were eligible for the water and sanitation program. Propensity score matching was used to identify the comparison segments that were most similar to treatment segments on observable variables thought to predict likelihood of being chosen for inclusion in a water and sanitation project area. When using the propensity score it is important to use variables that are strong predictors of selection into treatment. These variables usually consist of measures of the rules or the eligibility criteria in the program design.

[28] A detailed discussion of the covariates included in the matching model is provided in Appendix A. For the 2011 baseline survey, the list of segments that were expected to receive treatment was updated and the propensity scores were recalculated using the same statistical model used ex-ante. Accordingly, these segments do not necessarily overlap with the 2009 baseline sample.

The eligibility process was as follows:

> *Sixty-two municipalities in the Northern Zone, classified as either "Extreme Moderate Poverty" or "Extreme High Poverty" by the national poverty map, were invited to submit proposals for water and sanitation projects. To be considered eligible for the program, the proposals had to meet four criteria:*
>
> *a) the municipality had to be classified as either "Extreme Moderate Poverty" or "Extreme High Poverty",*
> *b) both the community and municipality had to be willing to make a financial commitment to the project,*
> *c) the community had to be organized and willing to work with the municipality, and*
> *d) the estimated cost of the project could not exceed $850 per beneficiary[29].*

In addition, some treatment segments were excluded before performing the matching if:

(1) The project associated with the treatment segment was very small (had fewer than 100 participants)
(2) The segment had fewer than 55 identified households (since it would be difficult to select more than 18 occupied households)
(3) The segment was also included in either the connectivity or northern zone multi-purpose household survey samples and elimination of the segment would not eliminate the entire project area from the survey sample.

Based on these criteria alternative specifications of the selection model were judged on the basis of their prediction rates and their ability to reduce bias between comparison and treatment groups. All specifications of the matching equation use variables that parallel the eligibility criteria used in the project selection stage, are predetermined and/or are correlated with the water availability in the segment. The final model used the following variables to predict whether or not a census segment would be a treatment segment (i.e. the probability of being a census segment that participates in the water and sanitation projects):

1) Indicators of location
    a. Average temperature
    b. Dummy variables for department
2) Characteristics of population and location
    a. Population of municipality in which segment is located
    b. Density of settlement
    c. Percent of surface area occupied by water bodies (an indicator of access to surface water)
    d. % of households in segment relying on private well for water
    e. Average household size (which is highly correlated with "rural area")
    f. Average number of household members who had emigrated (an indicator of external resources available to the community)
    g. % of households in segment with in-home business (an indicator of economic diversity and non-residential demand for water services)

[29] The costs per beneficiary ranged from $181 to $924 at compact close, thus some exceptions were made.

3) Indicators of inadequate water and/or sanitation
   a. % of households in segment relying on unimproved water source
   b. % of households in segment with piped water, but not receiving water every day
   c. % of households in segment with no improved sanitation facilities
   d. % of households in segment with composting latrines

"Annex 4: Matching Results" includes the logistic regression results as well as tables showing how the matching reduces bias, for the initial (2009) estimation and for the revised (2011) treatment assignment of segments. The estimation of the propensity score using the census data is presented in Table 17 and the test before and after matching in Table 20. Table 20 and Table 21 show that the PSM methodology is able to balance all the variables available in the census data with the treatment assignment of 2011[30].

Assumptions of the PSM and Balancing

The main assumptions of the PSM methodology to calculate the average treatment on the treated are [Heckman, Ichimura, and Todd (1998)]:

(1) Conditional on a set of covariates or conditional on the propensity score treatment status is mean independent of the outcome of interest,

$$E[Y_0|P(Z), T = 1] = E[Y_0|P(Z), T = 0] = E[Y_0|P(Z)]$$

(2) That the propensity score is bounded away from one, to allow us to find appropriate matches for each treated unit. This is what is called the 'common support' requirement in the literature.

$$P(T = 1|Z) < 1$$

Below we present the estimated propensity score distribution before and after matching and direct the reader to the baseline report for further evidence of the comparability between the treatment and control groups. Figure 3 shows the estimated propensity scores for the full sample of segments. On Figure 4 we can see the initial distribution of the propensity scores in 2009 and for 2011 in Figure 5 for the matched sample of segments. In Figure 4 we can see the considerable overlap of the probabilities of treatment for the selected sample after selecting the comparison segments using nearest neighbor matching for the the initial 2009 survey. There remains a region with little overlap to the right of the distribution which is wider in Figure 5, with the recalculated propensity scores that were used to select the 2011 baseline sample; these propensity scores reflect the treatment status changes in some segments and the selected comparison segments. This is due to the changes in the projects after the sample was selected and the limited the number of replacement segments that were feasible to map before the resampling for the 2011 baseline survey. The propensity scores after matching in both cases are very similar and have a wide overlap in their support; all comparison segments were selected from the region of common support.

[30] Table 14, Table 15 present the same figures, with the treatment assignment of 2009

FIGURE 3 DISTRIBUTION OF PROPENSITY SCORES BEFORE MATCHING

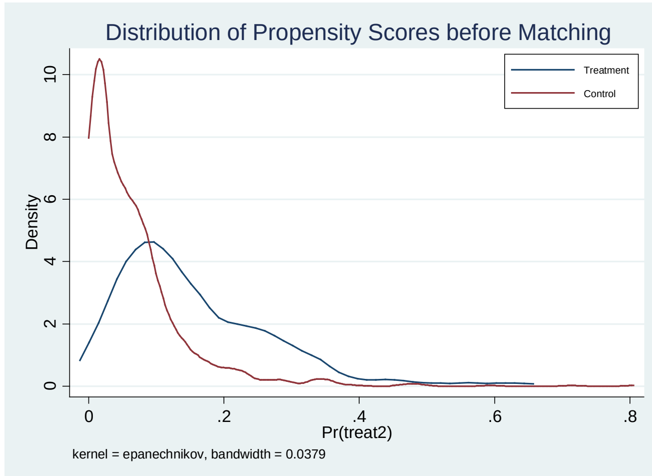


FIGURE 4 DISTRIBUTION OF PROPENSITY SCORES AFTER MATCHING 2009

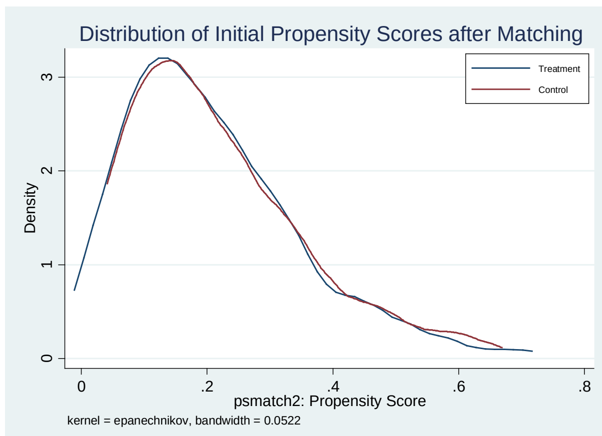
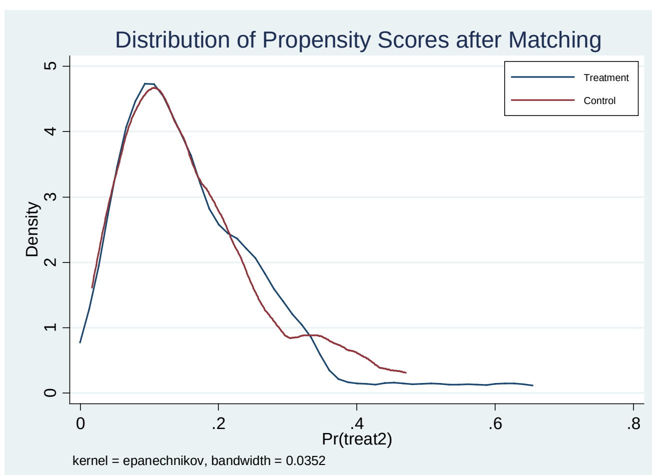


FIGURE 5 DISTRIBUTION OF PROPENSITY SCORES AFTER MATCHING 2011



4.4. Study Population

Sixty-two municipalities in the Northern Zone, classified as either "Extrema Pobreza Moderada" or "Extrema Pobreza Alta" (extreme moderate poverty or extreme high poverty, respectively) by the national poverty map[31], were invited to submit proposals for water and sanitation projects. To be considered eligible for the program, the proposals had to meet four criteria: (1) the municipality had to be eligible to participate, meaning there were classified as high or moderate extreme poverty; (2) both the community and municipality had to be willing to make a financial/labor commitment to the project, (3) the community had to be organized and willing to work with the municipality, and (4) the estimated cost of the project could not exceed $850 per beneficiary[32]. After projects that did not meet the eligibility criteria were excluded, a list of 68 projects remained. These were cleared to enter the feasibility stage. Comparisons segments were selected from non-beneficiary segments that where eligible to participate taking into account the poverty map, an proxies for financial capacity of the municipality and community involvement where included in the propensity score estimation.

As implementation of the first round of feasibility studies progressed, a new set of challenges to the evaluation surfaced: the cost of some projects was higher than expected, the water sources were not viable and/or the number of participants was lower than expected. These factors resulted in some projects being dropped, while others were expanded. The sampling frame was adjusted accordingly to be able to correctly identify treatment segments in the sample, maintain the power of the design and to measure the outcome indicators prior to the finalization of the projects.

In the original economic analysis of the water and sanitation component of the Compact, it was assumed that all water projects would involve providing water to households that did not previously have improved services. However, the project application, selection process, and feasibility studies generated a more diverse set of projects. Twenty-five of the projects installed water in communities that did not previously have improved water services. The remaining twenty projects extend an existing water system to additional households, improve an existing system, or both improve and extend an existing deficient system.

In the next sections we present the initial power calculation and the revised power calculations given the changes in the project that occurred after the initial power analysis and using the baseline data and follow up data. In summary, the power analysis shows that the survey is well powered to detect an effect as low as 8% percent in income/consumption, which is below the 15 percent income increase target expected by MCC.

[31] A household is in extreme poverty if their income is not sufficient to cover the cost per-capita of the basic food bundle. The classification of extreme poverty in severe, high, moderate and low is obtained using a k-means cluster analysis with the proportion of households in extreme poverty and the rates of stunting in each municipality. [FISDL (2005)]

[32] See " Annex 1: List of eligible municipalities" and "Annex 2: Project Selection Criteria"

4.5. Power calculations and sample size requirements

It is of vital importance in impact evaluation studies to address issues of power and sample size at the design stage of the study. With that in mind, in the initial design we used 2007 EHPM[33] (Multipurpose household Survey) and census data to construct our sample given the results of the PSM. The initial power analysis was conducted under the assumptions that the minimum detectable impact/effect should be a 10% increase in household expenditure, with 80% power at the 5% confidence level. It concluded that our survey needed a sample size of at least 3,168; with 88 control and 88 treatment segments, each with 18 households.

In subsection 4.5.1, we describe the original sample design, based on the EHPM 2007 calculations. Subsequently, Subsection 4.5.2 provides updated power calculations using data from the baseline survey collected in 2011 and the 2012 follow-up. In summary, the results show that the survey is powered to detect and effect as low as 8 percent change in income/consumption, which is below the 15 percent income increase expected by MCC. In addition we note that, the panel structure of the design will decrease the minimum detectable effects in our sample and that for some indicators, especially the coping cost the minimum detectable effects will be smaller.

4.5.1. Sample design and power for 2009 baseline

As described above, the impact evaluation will examine change in numerous indicators. Here we compute the sample size required to detect changes in expenditure.[34]

Indicators of incomes or expenditures are continuous. For continuous outcomes, the sample size required is given by [adapted from Purdon (2002)]:

[1]

$$C = \frac{(1 + (m-1)\rho)\left(z_{\alpha/2} + z_{\beta}\right)^2}{2\Delta^2 rm(1-a)a}\sigma_y^2$$

Where

ρ is the intracluster correlation (ICC)

m is the average number of subjects per cluster.

C is the number of clusters per condition

a represents the proportion of the total sample that will be allocated to the intervention group[35];

---

[33] Encuesta de Hogares de Propósitos Múltiples

[34] The sample size calculations have been done based on one primary indicator of household welfare (household expenditure); thus, there will be more uncertainty about whether we will have the power to detect changes in other, complementary indicators.

[35] Given we are using one-to-one matching and assuming the same cluster size, a must be equal to 0.5 so that

$$C = \frac{2(1+(m-1)\rho)\left(z_{\alpha/2}+z_{\beta}\right)^2}{rm\Delta^2}\sigma_y^2$$

σ_y^2 is the baseline variance of the continuous indicator

α is the type 1 error; the significance level to be used in the statistical tests;

β is the type 2 error, 1- β is the power of the study;

r is the response rate; and

Δ is the difference to be detected.

The effect of ICC on the sample size arises from design effects attributable to factors shared by observational units, thereby lowering the total amount of information available for statistical analysis and increasing the required sample size. By increasing the number of communities in the sample and reducing the number of households sampled per community (m), we can increase the amount of independent data. This will then lower the variance inflation factor, which is dependent on the ICC (ρ).

According to the 2007 EHPM household survey, average monthly household income for rural households and the monthly household expenditures and income were:

Mean Monthly Household Income (2007)	$284
St. Dev. of Mean Household Income	313
Mean Monthly Household Expenditure (2007)	$239
St. Dev. of Mean Household Expenditure	170

An estimate of the ICC was obtained from analyzing the 2007 household survey data and observed at the department level. Table 4 shows alternative ICC estimates that range from 0.0147 to 0.0396. Given that the upper estimates in this range are expected to be a significant subset of our sample, we used ρ=0.03 in our initial calculations.

TABLE 4 ESTIMATES OF THE INTRA-CLUSTER CORRELATION OF INCOME AND EXPENDITURE IN EHPM 2007

Description	ICC estimates
Monthly household **incomes**, rural households in departments with the majority of projects	0.0187
Monthly household **expenditures**, rural households in departments with the majority of projects	0.0147
Monthly household **incomes**, rural households without access to private taps in departments with the majority of projects	0.0262
Monthly household **expenditures**, rural households without access to private taps in departments with the majority of projects	0.0396

Using these data and equation [1], the sample size necessary to detect a 10% increase in monthly household income would be about 5,900 and the sample size necessary to detect a 10% change in monthly household expenditures would be 2,473. The calculations assume an α equal to 10%, β equal to 20%, and 20 observations per cluster.

We recommended focusing on household expenditures, as a proxy for household income in this study. Assuming the highest level of intra-cluster correlation that appears in the table above: $\rho=0.04$. We also assume a 10% attrition in each survey round (10% in the first follow up survey round in year 4, and an additional 10% in the last survey round) and add a contingency of 10% to each sample to account for possible changes in project design, problems with non-response or non-participants in treatment segments, or the loss of some clusters due to unforeseen circumstances.

Table 5 below shows the sample size required to detect a 10% change in expenditures under these assumptions. Consideration must be given to statistical, logistical, and financial matters in choosing how many clusters to include in the study. Statistically, it is usually advantageous to maximize the number of clusters as power is most directly affected by this number. At the same time, since we are going to compare community-level averages in treatment and control communities, a sufficient number of observations per cluster must be interviewed to ensure a representative sample is included for each cluster. Also, logistically and financially, it is usually more difficult and more expensive to recruit and survey more clusters than it is to interview more observations per cluster.

TABLE 5 SAMPLE SIZE REQUIRED TO DETECT 10% EXPENDITURE EFFECT

Required sample size including 10% loss to follow up and 10% contingency[36]	Number of Communities	Obs. Per cluster (with 10% contingency and 10% loss to follow up)
2,745	189	15
2,973	164	18
3,202	147	22
3,431	135	25
3,660	126	29
In a General Population Sample (ICC=0.04)		

We ultimately adopted 18 observations per cluster and 164 communities, while adding an additional contingency -- 6 extra treatment segments and 6 extra control segments, for a total of 216 additional households in order to provide extra cushion for the loss of projects during implementation or inaccuracies in the sample frame. The final sample size for recommended for the study was 3,168, with 88 comparison and 88 treatment segments, each with 18 households.

4.5.1. Selection of households

The final step in the selection of the sample was the selection of households. DIGESTYC staff visited all chosen treatment and comparison segments to map the segment and create a sample frame of occupied houses. Only occupied buildings that served as either a residence only or as both a business and a residence were retained in the sample frame. Eligible buildings were then listed according to the DIGESTYC standard numbering system (working from one side of the segment to the other), and grouped into blocks of four. Systematic sampling was used to draw six blocks of four (consecutive) residences each, for a total sample of 24 households.

In each block, the first three residences (in numerical order according to the DIGESTYC code) were visited by the enumerators on the day of the survey. If one of these three houses was found not to be eligible for the survey *because of a problem with the sample frame*, then the enumerator visited the fourth house as a replacement. Only one replacement house was available for each block.

If more than one household was found to live in a particular residence that had been selected into the sample, the primary household (the household of the owner or the household that had lived longest in the home) was chosen as the household to be interviewed. This was done because we want to have the best possible chance of finding interviewed households again at the time of next survey.

In some cases, occupied houses were stripped from the sample frame before the selection was made. This was done when the sample frame produced by DIGESTYC indicated that the treatment segment included a large number of households who did not live in the (anticipated) project area

[36] Note that the required sample size in the table does not exactly equal the number of communities or the observations per cluster because of how the 10% contingencies were calculated and added

(according to the project description available at the time of the sampling). In order to increase the probability that the sample in "treatment segments" would comprise treated households, those households outside of the anticipated treatment area were eliminated.

In what follows we revise, these power calculations for the sample that was drawn in the 2011 baseline accounting for the changes in the projects from 2009 to 2011.

4.5.1. Updated Power Calculations

In this section we review the initial sample size recommendations and update the parameters used for power analysis using the baseline data, to emulate the calculations based on the EHPM and then proceed to account for the overtime variation in outcomes given that we have two rounds of data from the impact evaluation survey.

In updating the power calculations, we use the 2011 baseline survey to calculate the intra-cluster correlation and variance needed to calculate the power of the survey; we then follow this with a power analysis focusing on the minimum detectable effects[37] (MDE) for the sample size we have in the survey. The former serves to contrast the initial recommendations based on the EHPM and what the recommendations would be using the baseline survey data. The latter serves as a post-hoc power analysis to revise our expectations regarding the precision and power of any impacts that are detected. These revisions take into account the changes in the sample due to changes in the project areas. The general purpose is to know what differences we are able to detect in the final impact analysis and to weigh any gains or losses that might have occurred since the initial sample design.

The calculations are made under the following parameters, assumptions, and estimations:

TABLE 6 POWER CALCULATION PARAMETERS FROM 2011 BASELINE SURVEY

Mean Monthly Household Expenditure	**230.5**
Std. Dev. of Mean Household Expenditure	110.7
Minimum Detectable Effect	10%
Intra-class Correlation	0.08
Average Cluster Size	23.8

In Table 6, it should be noted that the average number of households per cluster (segment) was well above the initial recommendations, with 23.8 households per segment. Also, the intra-cluster correlations is higher than that observed in the EHPM 2007; consequently, having a smaller number of households per cluster would have been more efficient, since a higher intra-cluster correlation implies that it is more efficient to have more clusters and reduce their size for a given sample size. The effect of this high intra-cluster correlation was diminished by the lower variance of monthly consumption found in the baseline survey.

In Table 7, we calculate the sample size needed given the observed mean and variances in the 2011 baseline survey and present two scenarios: one using the standard 80% power with a 5% and 10% confidence level in Table 7. Table 8 shows the realized sample for the baseline and the follow-up.

[37] Intuitively, the MDE of a study is the smallest true effect that can be detected with acceptable certainty

Comparing these two tables we can see that the sample is well above the recommendations based on the baseline estimates. The attrition has also been lower than what we adjusted for in the initial calculations. Figure 6 shows the number of clusters needed to detect different effects.

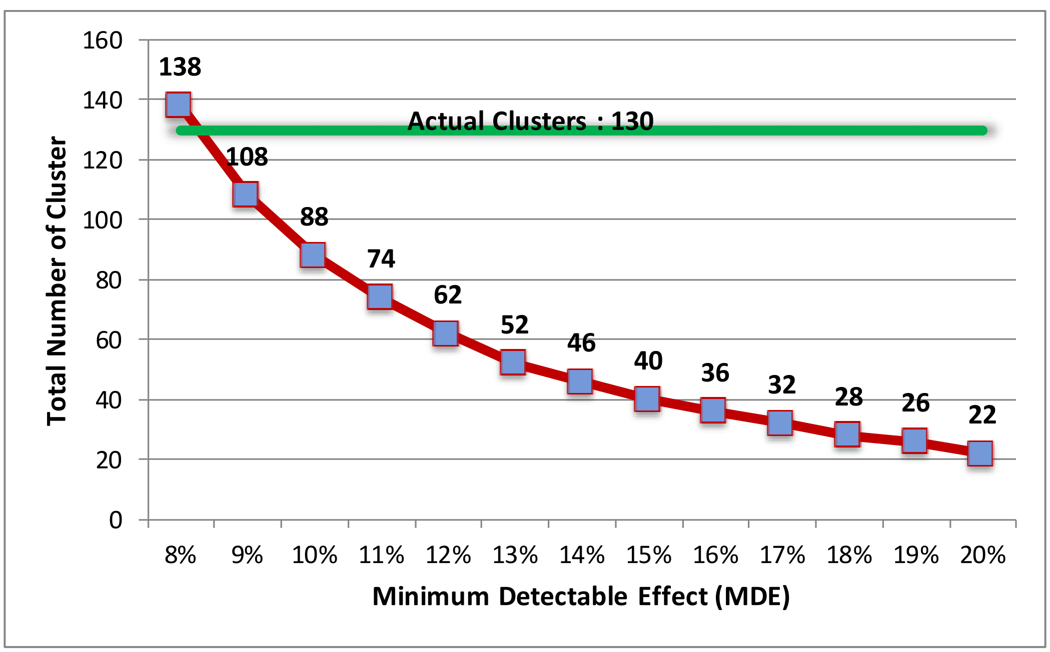
TABLE 7 SAMPLE SIZE REQUIREMENTS GIVEN 2011 BASELINE

					Adjusting 10% Clusters for attrition		
Alpha	**Beta**	**Clusters per condition**	**Total Clusters**	**Total Households**	**Clusters per condition**	**Total Clusters**	**Total Households**
5%	20%	44	88	2,092	48	96	2,282
10%	20%	35	70	1,664	38	76	1,806

TABLE 8 SAMPLE SIZE AT BASELINE AND FOLLOW-UP

	2011		2012		
	Households	**Clusters**	**Households**	**Clusters**	**Attrition**
Control	1,637	65	1,560	65	4.7%
Treatment	1,647	65	1,529	65	7.2%
Total	3,284	130	3,089	130	5.9%

FIGURE 6 CLUSTER VS. MDE WITH BASELINE ESTIMATES

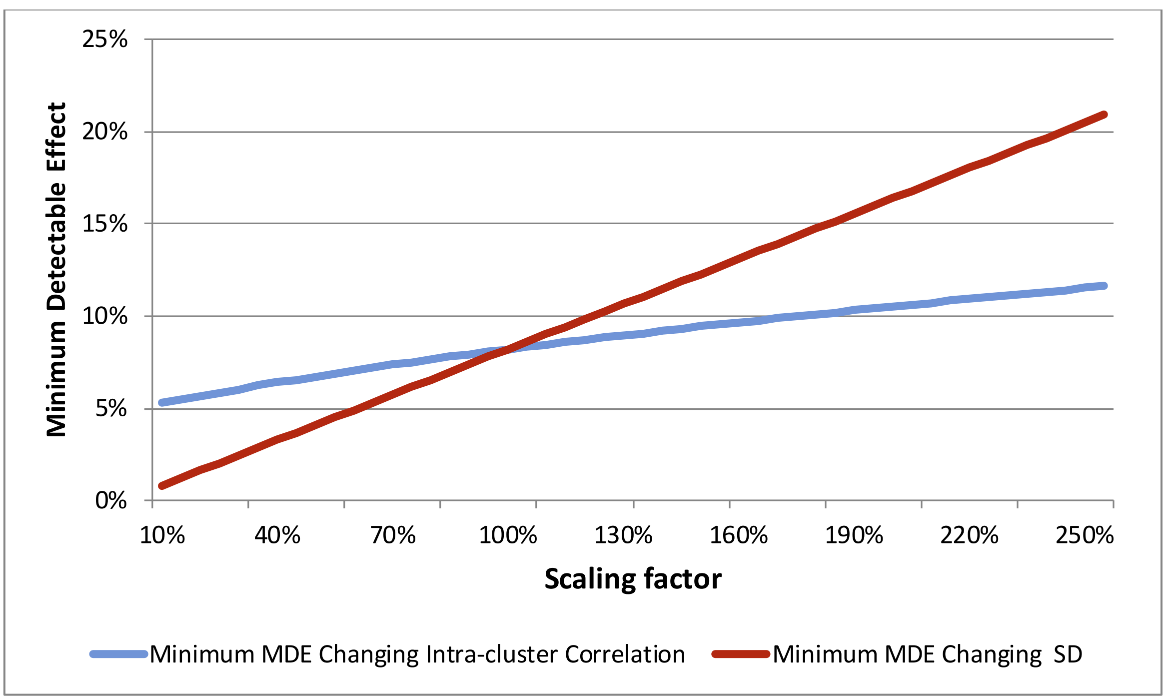


To gauge the effect of these deviations on the power of the proposed test, we calculate the minimum detectable effects for the realized sample using a post-treatment comparison methodology. The equation for the MDE takes the form (from equation 1):

$$\hat{\Delta} = \sqrt{\frac{2(1+(m-1)\rho)\left(z_{\alpha/2}+z_{\beta}\right)^2}{Cm}\sigma_y^2}$$

The minimum detectable effect (MDE) is below the planned 10%. Given the changes in the sample, we can detect a 8.19% impact in expenditure at the 5% confidence level and with 80% power. However, given the significant changes we found in the intra-cluster correlation and variance of the baseline vis-à-vis the EHPM, it is important to gauge how the MDE changes when we change these variables. Figure 7 shows this exercise. In this figure we scale the observed standard deviation (green) and the intracluster at baseline by 10% to 255% (11 to 282 the standard deviation and .008 to 0.209 the intracluster correlation). The MDE is more sensible to changes in the standard deviation, with a 20 percent increase in the standard deviation bringing the MDE above the 10%. In the case of the intracluster correlation, an increase of over 75% brings the MDE above the 10%, but at a slower pace.

FIGURE 7 MDE UNDER CHANGES OF INTRA-CLUSTER CORRELATION AND VARIANCE



Note that the MDEs above are calculated with baseline data and under the assumption of the post-treatment test. We will see some gains in precision and power by controlling for covariates and the multiple rounds of data available. These two aspects of the design should improve the power of the study and the diminished the effect of attrition in the effective sample size at end-line. Next we report the power calculations based on the two rounds (2011-2012) available at the time of writing using the following equations adapted from Murray (1998) (we call these MDEs with panel adjustment):

$$\hat{\Delta} = \sqrt{\hat{\sigma}_{\Delta}^{2}\left(z_{\alpha/2} + z_{\beta}\right)^{2}}$$

$$\hat{\sigma}^{2}{}_{\Delta} = 4\left[\frac{\hat{\sigma}^{2}{}_{m}\hat{\theta}_{m}\hat{\psi}_{m}\left(1 - \hat{r}_{yy(m)}\right) + m\hat{\theta}_{C}\hat{\psi}_{C}\hat{\sigma}^{2}{}_{C}\left(1 - \hat{r}_{yy(C)}\right)}{Cm}\right]$$

$$\hat{\theta}_{d} = \frac{\hat{\sigma}^{2}{}_{d}|X}{\hat{\sigma}^{2}{}_{d}} \; for\ d = m, C$$

$$\hat{\psi}_{d} = \frac{\hat{\sigma}^{2}{}_{d}|P(Z)}{\hat{\sigma}^{2}{}_{d}} \; for\ d = m, C$$

$\hat{r}_{yy(C)}$: overtime correlation within clusters

$\hat{r}_{yy(m)}$: overtime correlation within households

$\hat{\theta}_{m}$: ratio of the X covariates-adjusted variance to the unadjusted variance attributable to households

$\hat{\theta}_{C}$: ratio of the X covariates-adjusted variance to the unadjusted variance attributable to clusters

$\hat{\psi}_{m}$: ratio of the Z-matched variance to the unmatched variance attributable to households

$\hat{\psi}_{C}$: ratio of the Z-matched variance to the unmatched variance attributable to clusters

So that the MDE, accounting for matching and covariate adjustment can be written as

$$\hat{\Delta} = \sqrt{4\left[\frac{\hat{\sigma}^{2}{}_{m}\hat{\theta}_{m}\hat{\psi}_{m}\left(1 - \hat{r}_{yy(m)}\right) + m\hat{\theta}_{C}\hat{\psi}_{C}\hat{\sigma}^{2}{}_{C}\left(1 - \hat{r}_{yy(C)}\right)}{Cm}\right]\left(z_{\alpha/2} + z_{\beta}\right)^{2}}$$

Table 9 shows the power analysis results using the two rounds available and assuming one impact estimate for the full post-treatment period. The results show a gain in precision, given from the panel. Finally we present Figure 8 with the range of possible effect sizes that can be detected given our sample for different values of the ratio to adjust for matching and for the covariates. We present the MDE's as a function of $\hat{\theta}_{m}\hat{\psi}_{m}$ (in red-attributable to households) and $\hat{\theta}_{C}\hat{\psi}_{C}$ (in green-attributable to clusters). In general, the $\hat{\theta}_{m}\hat{\psi}_{m}$ tends to remain stable or decline, while $\hat{\theta}_{C}\hat{\psi}_{C}$ could increase or decrease. The figure shows that the MDE remains below the expected 15%, for a wide range of scenarios; the matching and covariate adjustments will entail some gains in precision.

TABLE 9 MDE WITH PANEL ADJUSTMENT: 5% CONFIDENCE, 2 ROUNDS POST TREATMENT

α	β	Mean expenditure at Baseline ($)	Households per Cluster	Clusters per condition	Within cluster Correlation across t:	Within household Correlation across t:	Variance of Impact estimate	Minimum Detectable Difference ($)	Minimum Detectable Difference (%)

					$r_{yy(g)}$	$r_{yy(m)}$			
5%	20%	230.53	23.8	65	0.55	0.07	41.89	18.20	7.9%
10%	20%	230.53	23.8	65	0.55	0.07	41.89	16.14	7.0%

FIGURE 8 PLAUSIBLE RANGES OF MDE GIVEN MATCHING AND COVARIATE RATIOS

4.5.2. Implications for measuring other impacts

Since the primary objective of the research design is to measure changes in expenditure and coping costs in a rigorous way and the sample size was design with the expenditure variable. We use the baseline data to calculate the MDE's for other important measures of income, coping costs, and water related behavior; some of which was not possible to calculate using the EHPM. Table 10 shows the MDE's with and without the panel adjustment as before. The results show that the realized sample is well powered to detect sensible changes in these variables. For example, we can detect an impact of 8 percentage points in the poverty incidence, 6 percentage points in the probability of carrying water from outside, 0.63 hours per week difference in the time spent carrying water in the households, etc.

TABLE 10 MDE FOR OTHER IMPACT INDICATORS

	Mean at Baseline	Variance	Within household Correlation across time	Within cluster Correlation across time	Variance of Impact estimate	MDE WITH panel adjustment	MDE WITHOUT panel adjustment
Log Annual Consumption	7.80	0.30	0.09	1.00	0.001	**0.08**	**0.10**
Annual Gross income (Thousands $)	2.83	9.67	0.06	0.97	0.035	**0.53**	**0.48**
Log Annual Gross Income	7.50	1.08	0.09	1.00	0.003	**0.15**	**0.17**
Poverty incidence at 2011 poverty line	0.67	0.22	0.07	1.00	0.001	**0.06**	**0.08**

Monthly expenses in water	7.39	161.22	0.13	0.93	0.293	**1.52**	**2.36**
Log Monthly expenses in water	1.12	1.92	0.25	0.99	0.003	**0.16**	**0.33**
Probability of carrying water	0.32	0.22	0.20	0.99	0.000	**0.06**	**0.11**
Probability of doing laundry outside	0.24	0.18	0.18	1.00	0.000	**0.05**	**0.10**
Time Carrying water (Hrs per Wk.)	1.99	26.76	0.10	0.94	0.050	**0.63**	**1.01**
Time Laundry (Hrs per Wk.)	2.21	38.38	0.11	0.97	0.078	**0.79**	**1.24**
Water Consumption (Thousands of Lts. per Wk.)	2.01	6.43	0.14	1.00	0.011	**0.30**	**0.52**
Average time to all water sources (Minutes)	8.05	100.84	0.15	1.00	0.188	**1.22**	**2.12**
Chlorine above 0.3 mg/Lt.	0.10	0.09	0.22	0.97	0.000	**0.06**	**0.07**
Tap water is principal source for Drinking	0.39	0.24	0.41	0.99	0.000	**0.06**	**0.16**
Tap water is principal source for Cooking	0.41	0.24	0.43	0.99	0.001	**0.07**	**0.17**
Tap water is principal source for Washing	0.40	0.24	0.41	0.99	0.001	**0.07**	**0.17**

Note: Calculations assume 5% type one error and 80% power. The average size of the clusters is 23.8 with 65 clusters per condition

4.5.3. Implications for measuring health impacts

The final sample is not designed to measure, on its own, the health impacts caused by the WSS. The primary objective of the research design is to measure changes in expenditure and coping costs in a rigorous way. To address health in the evaluation, we propose to combine at least two sources of information to estimate health impacts (albeit not with the precision and power possible by adding a separate sample of households with children under three years old).

First, as the general population sample described above is randomly drawn, we would expect that about 22% of households in the sample (roughly 600-700 households) will be households with children under three (based on data on the number of households with children under three in the census data). The sample of children under age three is important for measuring changes in the diarrhea rate because the vast majority of diarrhea cases occur in children under five. We are adopting a panel survey design, with two years between the first and last survey. To see changes in the diarrheal rate, we need to identify households that will have children under age five at the time of both the first and last survey. Households that in 2011 have children under three will still have children under five in 2013. Indeed, in the 2011 baseline survey, we had 693 households with children under the age of three in the sample.

These households with children under three would play two roles in our estimates of the health effects of the WSS interventions. Most importantly, the small sample of households with children under three would be our primary source of information about the time and monetary costs of illness – how much do these households spend on a treating diarrhea and how much productive time is lost? Both pieces of information are important inputs into the ERR for the water and sanitation projects. The sub-sample of households with small children may also be large enough to detect a very large decrease in diarrhea rates among this population. With a general population sample of about 3,400 households, for example, there should be a sufficient number of households with children under three years (expected n=755) to detect a diarrhea rate decrease of between 40-50% during the dry season (Table 11). It is probably unrealistic to expect to see reductions of this magnitude in El Salvador given the progress that has been made in reducing diarrhea rates in recent years. For example we only find 190 cases of diarrhea in 158 households for a 15-day reference period; that is below 1.3% incidence. Nonetheless, if there were to be a large change, it is likely that we would be able to detect it with this small sample of households under age three. It is important, however, not to expect too much from this small sample of children under three. Not only is the effect size that this sample could detect large, these effects would also be measured with a sample of at most six households with children under three per census segment. Since these are likely to be only a small fraction of the households with children under three in the segment, we could not claim that the households in a community are representative of the effects in that segment.

TABLE 11 SAMPLE SIZES REQUIRED TO DETECT CHANGES IN DIARRHEA RATES IN HOUSEHOLDS

Effect Size	90% of April-May Average Estimated 2007 Diarrhea Prevalence in under 5s =17.4%*	90% of May - July Average Estimated 2007 Diarrhea Prevalence in under 5s = 26.5%*
35%	1,233	833
40%	919	623
50%	556	378

***Percentage of population under 5 who experience diarrhea at least once in this time period. Assuming about six observations per segment and 76 segments**

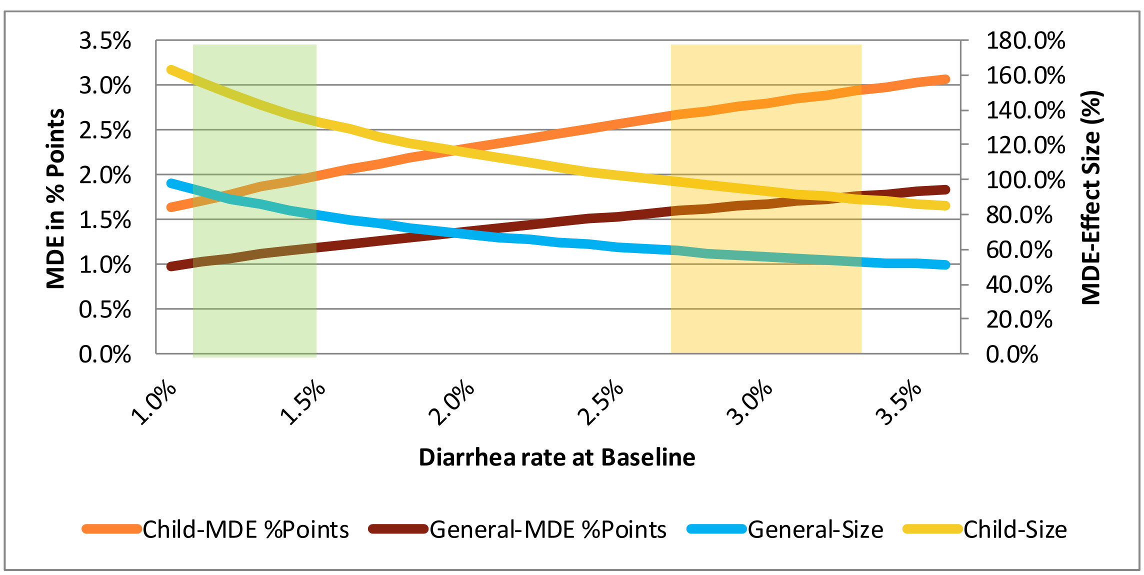
In the baseline sample, the average diarrhea incidence was 1.3% of all individuals and 3% of children under five. We calculate the MDE using [Donner, Birkett, and Buck (1981) Diggle, Liang, and Zeger (1994) Donner (1998) Purdon (2002)],

$$\hat{\Delta} = \sqrt{\frac{2(1+(m-1)\rho)\left(z_{\alpha/2}\sqrt{\bar{\pi}(1-\bar{\pi})} + z_{\beta}\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}\right)^2}{Cm}}$$

Where all variables are defined as before and $\bar{\pi} = \frac{\pi_1 + \pi_2}{2}$, and π_t is the proportion of households/individuals[38] reporting diarrhea in period t.

Figure 9 shows the MDE (in percentage points and percent of baseline) that we can detect given the baseline sample for different baseline diarrhea rates.

FIGURE 9 MDE FOR DIARRHEA IN GENERAL AND CHILD POPULATION



[38] The proportions are similar for households and individuals.

The range shaded in green, shows the area around the baseline estimate for the percentage of people reporting diarrhea in the full sample. The figure shows (blue and red lines), that a 1 percentage point difference in this region, implies very large effect sizes of around 60%. In the case of children under 5, (orange shaded) the MDE size is even larger, since even though the diarrhea rate is larger, the number of observations per cluster is much smaller (6 household with children under 5 per cluster vs. 24 households per cluster in the general sample). On this front, effects on diarrhea will be difficult to detect, given the low incidence of diarrhea in the sample.

Second, we propose to leverage existing epidemiological research that measures the health impacts associated with exposure to water and sanitation interventions to estimate the health effects likely to have been realized as a result of the Compact's WSS investments. There have been several recent studies summarizing the epidemiological research on the health effects of water and sanitation. For example, Fewtrell et al. (2005) report a meta-analysis of studies evaluating the health impacts of water and sanitation interventions. Our evaluation will measure the changes in water and sanitation outcomes (e.g., access and use of improved source) in the treatment and comparison areas through the general population panel survey, and then use existing epidemiological study results to calculate how these changes in outcomes are likely to influence health in the El Salvador case. As a test of this approach, information from the sub-sample of households with small children will be used to help assess the applicability of the existing epidemiological studies to the El Salvador case.

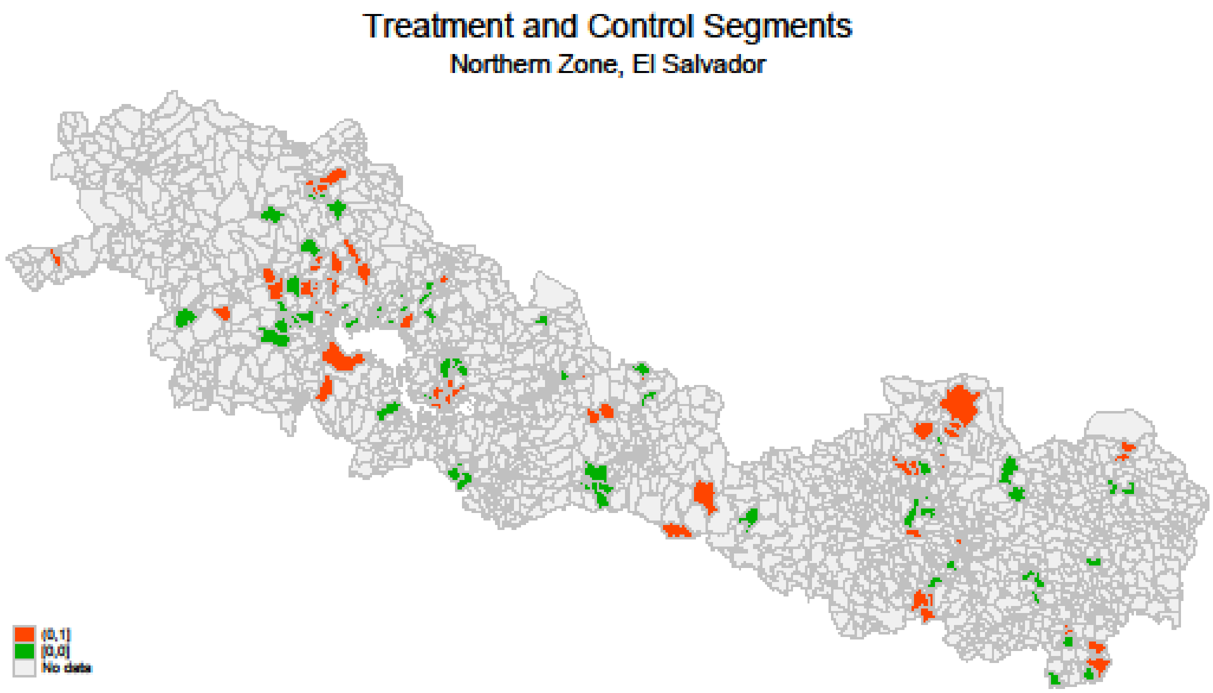
In addition to these two options, we will try to measure changes in diarrhea rates in comparison and treatment areas by studying information from the census kept at public health posts (municipal or sub-municipal level). These records are kept for five years and include both the purpose of the visit as well as the residence of the patient. By examining a sample of census records in the clinics associated with our study population, we will compare the number of diarrhea-related visits by residents of treatment and control areas during a designated period in the baseline year, and in at least one follow up year. This information could be an important complement to the information about diarrhea rates gathered in the sub- sample of households under three and to the existing health department information on diarrhea rates at the municipal level. As records are kept for five years, this information will be gathered during the third year of the panel survey with a complementary survey instrument to characterize the health post, infrastructure and capabilities.

4.5.4. Geographical Distribution of Sample Census Segments

The geographical distribution of treatment and control segments are presented in Figure 10 for the entire Northern Zone of El Salvador. Treatment segments (in orange) comprise 65 segments with 1,637 households and control segments (in green) comprise 65 segments[39] with 1,647 households.

[39] Most segments have between 24 and 27 observations.

FIGURE 10 GEOGRAPHICAL DISTRIBUTION OF TREATMENT AND CONTROL SEGMENTS



4.6.Timeframe

The timing of the household survey activities are planned as follows:

- Three household and community survey rounds[40] (in March – May of years 2011, 2012 and 2013)
- Roughly 3,300 surveys per round in 130 census segments
- The content and length of the survey will be roughly the same in each year.
- In all three years, enumerators will do tests of residual chlorine levels in the drinking water of all interviewed households.
- Maps of sampled census segments including inhabited properties will be prepared.

Parallel to the household survey, a private firm will do testing for bacterial contamination in drinking water (at the household level) and at a selection of sources in the community[41].

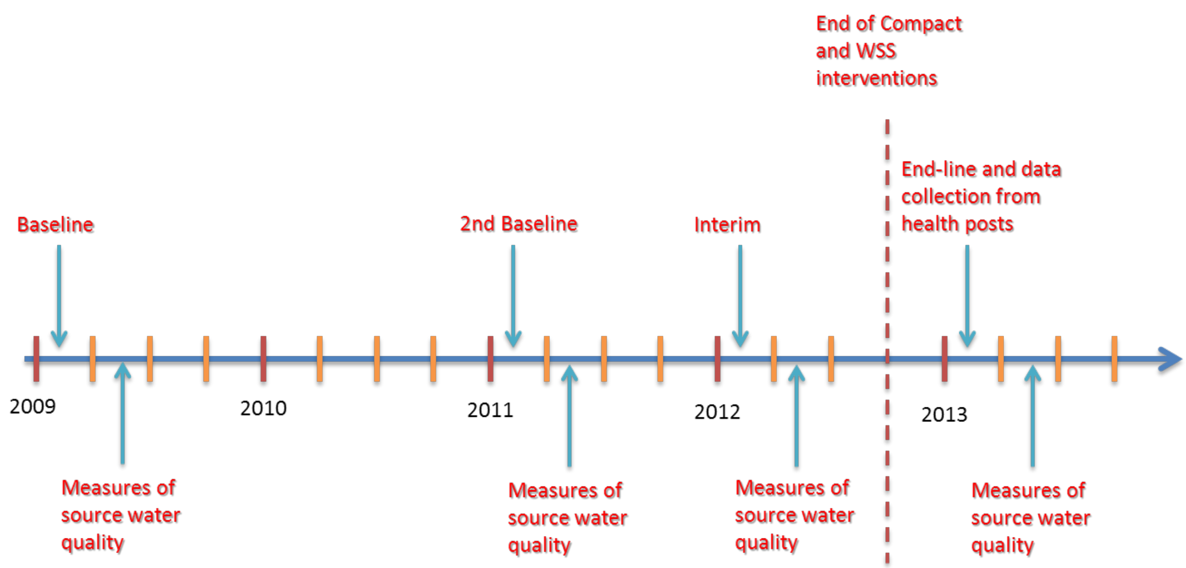
The timing of the community and health posts survey activities are planned as follows:

- Community surveys will be implemented at the same time as each of the three survey rounds, in each of the census segments included in the survey sample. These will involve surveys of a community leader about characteristics of the community and prices and services in the community. The survey will also include questions regarding status of water sources and WSS interventions.
- The survey in health posts will be done in 2013.

[40] As mentioned before we also have a 2009 baseline survey with partial overlap with the 2011-2013 surveys
[41] See "Annex 5: Water quality testing" for the details of the water quality testing.

An analysis of the baseline survey was submitted in February 2012 and an interim report in January 2013. The final report analyzing the end-line survey, water quality measurements, and any data collected from the Ministry of Health is scheduled for April of 2014. The final report will include the 3 waves of the survey and the effects for the follow-up and the endline.

FIGURE 11 TIMELINE OF DATA COLLECTION ACTIVITIES



4.6.1. Justification for proposed exposure period to treatment

In September 2012 the Compact closed and 45 projects have been completed. However, by the time of the first follow-up (March-May 2012) not all projects had been finalized, thus some treatment households were not receiving water from the improved systems by the time the survey was administered. The households that were receiving water from the improved water systems had as little as 1 month connected to the new systems. By the time of the end-line in March-April 2013 all beneficiary households were connected to the improved systems and will have been exposed to the treatment for different lengths of time; some might have as much as 2 years and others as little as 6 months.

Given this varied exposure time it is important to have multiple data points to estimate the full term and short term impacts of the projects. It permits us to explore the evolution of the main impact indicators and allow the beneficiaries to learn and change their behavior which would allow for the expected impacts to manifest in the data at later rounds.

By collecting data before and after program implementation, we can apply the 'difference-in-difference' estimator and the deviations from means or fixed effects estimator. These estimators measure the treatment effect as the difference between the changes in indicators before and after the project among treatment recipients, on the one hand, and the changes in indicators before and after the project among control units, on the other. In the case of the fixed effects the changes are deviations from the mean. These estimators help control for residual confounding due to imperfect matches and selection bias from time-invariant unobservable factors which differ between treatment and control communities and which may have an influence on the impact variables of interest.

Implementing a baseline and two follow-up surveys has a number of advantages over an evaluation design utilizing only two data points. The longer one waits to conduct the follow-up survey, the greater the risk that the measured impacts will be due (at least in part) to changes that are unrelated to the program under study and that households will forget key details about project implementation, such as how long they have been connected to the improved systems, their contribution in work hours or in monetary terms, and other aspects of community involvement that were part of the project. There is also the risk that some control communities will receive water and sanitation services from a different funding agency or program. Also, multiple data points increases the precision of the estimated impacts and ensures that biases due to the attrition in the panel are minimal given the yearly frequency of data collection. With a follow-up survey conducted both one and two years after the baseline we both minimize risk of contamination of comparisons and gain a view of the impacts after households members have had time to adjust their behavior .

4.7.Challenges for evaluating the impact of the investments

The Compact WSS poses some particular challenges for evaluating the impact of the investments.

- First, the program includes different types of projects, which will benefit households in different ways. For example, it is not obvious that projects that provide a new service option to a household without piped water service should be expected to have the same impact as projects that help households with connections get more out of a service they already have. For future policy discussions in El Salvador and within MCC, it could be very useful to understand the differential impacts of different types of projects. These more detailed research questions, however, would come at the cost of increasing the household survey sample size needed for the evaluation.
- Second, not all of the expected benefits of the water and sanitation program are relevant for all households. Whereas increases in income or reduction in water collection costs might be seen in all households, changes in school attendance rate relate primarily to households with school age children, and changes in morbidity and health expenditures from water and sanitation-related illnesses are expected to arise disproportionally in households with young children. This again has implications for sample design and sample size.
- Third, the benefits of water and sanitation investments have a seasonal character – households experience the benefits of service improvements in different ways at different times of year. Impacts on disease are most visible during the rainy season when diarrhea rates are highest. Impacts on coping costs (cash and time expenditure on water provisioning) are highest in dry season, when deficient pipe systems and shallow wells or streams are likely to experience water shortage problems.

We discussed these challenges and their implications for the research design and the sample size and concluded that we should give priority to the measurement of changes in household welfare and coping costs, over health impacts. There are two reasons for this. First, coping costs accounted for the bulk of the expected benefits of water supply interventions in the economic analysis underlying the program. Second, diarrhea rates in El Salvador have dropped substantially in recent years due to a variety of public health interventions. This makes measuring changes in rates more difficult and more costly, and also is likely to mean lower than expected benefits from health improvements in these projects. Because we will focus on changes in welfare and coping costs, benefits will be measured in the dry season only. Also, to keep down the sample size, we have decided to make the principle objective of the research the evaluation of the combined impact of the *full set* of water and sanitation projects. We will assess to what extent it will be possible to compare impacts across project types with the existing sample (albeit with less precision than for the water and sanitations projects as a whole).

5. Data Sources and Outcome Definitions

The impact evaluation uses three waves of household surveys and community surveys; and one wave of retrospective health post survey. Table 12 shows a list of the primary indicators that will be analyzed. In addition, Table 13 shows the definitions and targets of the primary indicators in the monitoring and evaluation (M&E) plan [MCC (2012)]. These indicators will be estimated using the panel household and community surveys.

TABLE 12 PRIMARY INDICATORS

Subject	Indicators
Consumption/ Expenditure	• Average total annual household expenditures
Income	• Average total household annual income • Average household income from business • Average household income and individual income (by gender and age) from wage labor
Productive time use	• Average time spent on income-earning activities, by household and by gender of the head of household • % of women / men / girls / boys engaged in wage labor • % of women / men / girls / boys working in a household business
Education	• School enrollment rate, for children ages 7 to 12, ages 6 to 18 by gender • Average days of school attendance during the 2 weeks prior to the survey, by gender
Health	• Diarrhea rate among children under 5, and among all individuals, during 2 weeks prior to the survey • % of households with adequately chlorinated water
Coping costs	• Average time spent collecting water, by household and by gender and age of household member • Average monthly expenditures on water, total and by source • Average monthly expenditures on treatment, maintenance, alternative sources and storage
Coverage and access	• % of households with improved water service at the home, in the community • % of households with latrine or other improved sanitation service at the home • Average distance household travels to water source
Water use	• % of households using improved water service for drinking and cooking • % of households that still using unimproved water service for drinking and cooking • % of household that use household latrine for sanitation needs

TABLE 13 M&E INDICATORS BASELINE AND TARGETS

Indicator	Definition	Baseline	Target
Increase in income of water and sanitation beneficiaries	Percentage increase in income of households receiving water and sanitation investments	0	15
Incidence of diarrhea	The percentage of individuals reported as having diarrhea in the two weeks preceding the survey	8.5	0
Reduction in days of school or work missed as a result of water-borne diseases (days per year per person)	Reduction of the number of days of school or work missed per year per person as a result of intestinal parasitism, diarrhea or infectious gastroenteritis per beneficiaries	0	-39
Cost of water	Average price of water per cubic meter paid by beneficiaries	1.68	0.43
Residential water consumption	The average water consumption in liters per person per day	69.42	177.1
Time collecting water (hours per week per household)	Hours per week per household spent collecting water by Project Activity households	4.58	0
Access to improved water supply	The percentage of households in the project area, whose main source of drinking water is a private piped connection (into dwelling or yard), public tap/standpipe, tube-well, protected dug well, protected spring or rainwater	79	83
Access to improved sanitation	The percentage of households in the project area who get access to and use an improved sanitation facility such as flush toilet to a piped sewer system, flush toilet to a septic tank, flush or pour flush toilet to a pit, composting toilet, ventilated improved pit latrine or pit latrine with slab and cover	82	83

5.1. Data Sources

The household level survey is administered in the departments of Cabañas, Chalatenango, Cuscatlán, La Unión, Morazán, San Miguel and Santa Ana. The survey is composed of a set of sections to characterize the water access situation of households, household demographics, consumption, income/productive activities and time allocation of women and children.

The community level survey includes 130 census segments representing 196 caseríos[42]. The information is obtained from interviews of key informants from the communities. Key informants include health workers/promoters, members of the water boards and other community leaders.

[42] Segments can include more than one caserío or include only a part of a caserío. Segments will be referred as such or as communities.

5.1.1. Quantitative – Household Survey

The primary source of data for the outcome and impact indicators will be the household panel survey. This survey will be administered to heads of household or adults older than 16 years old. Enumerators will give priority to interviewing the person in the household with the best knowledge of household expenditure and water use and collection. If the person is not present at the time of the first visit, enumerators will attempt to make an appointment and return again to interview the appropriate person, provided that this return visit is possible within the time that the survey team will be in the area. When possible, a second adult can also be included in the interview process, particularly for the questions related to work and agricultural output. The survey is designed to take between 1 and 1 ½ hours.

Table 14 describes the survey and some of the main indicators used to measure changes in wellbeing after improvement of water and sanitation services. The water access modules characterize the ways households obtained water at the time of the survey; it provides information on the types of systems the household use to obtain water, their perception of the risk of using the water, the cost of obtaining it and the availability/reliability of the water source. The household characteristics modules elicit information on the education of the household members and the characteristics/state of the infrastructure of the home. The labor modules provide information on the types of activities people engage in (agricultural and non-agricultural, salaried and non-salaried) as well as the income earned by the household. The labor, farm production, income, and inputs modules allows for the measurement of changes in income and household expenditures resulting from improved water and sanitation access. The time allocation module provides information about how household women and children allocate their time, which allows for measurement of the indicators related to hours spent working, taking care of children, carrying water and doing other household chores.

TABLE 14 HOUSEHOLD AND COMMUNITY SURVEY OUTLINE

Survey	Modules		Outcomes/Indicators
Household and Individual	Water access through informal systems, public and private taps, wells, natural sources, etc.		Quality, availability of water and coping costs of obtaining water
	Household Characteristics		Time spent collecting water
	Salaried, Independent, Agricultural Labor		
	Agricultural Production and Income		Changes in income and expenditures
	Time allocation		
	Health and Sanitation		Incidence of water related illnesses
Community	Prices and events in the community		
	Water, Sanitation, Electricity and services		Water Consumption, Reduction in days of school or work missed as a result of water-borne diseases

5.1.2. Quantitative and Qualitative - Community Survey

The community surveys gather information about the local economy; price levels for food, basic commodities, and water and sanitation -related expenditures; and about the history of public services in the area. These surveys will also include questions regarding the status of water sources and the water and sanitation project implemented through the Compact. These surveys will be administered to one or more community-based organization or community leader. Price information will be collected in stores, from vendors, or in interviews with knowledgeable community members. The goal of the surveys is to provide some context for the information gathered in the household surveys, to track community-level changes that may affect outcomes, and to reduce the required length of the household survey questionnaire.

6. Analysis Plan

This section spells out the econometric analysis that will be implemented for the DID methodology described in Section 4.3.

6.1. Statistical techniques

The DID estimator measures the treatment effect as the difference between the change in an indicator before and after the program among treatment recipients and the change in the indicator before and after the program among control units. This approach removes the effects of any unobserved fixed factors that differ between participants and non-participants, if those have a linear, additive impact on outcomes. However, the results may be confounded by other changes between the time of the baseline and the follow up surveys that differentially influence changes in outcomes between participants and non-participants (e.g., changes in access to other programs).

We employ regression-adjusted DID estimation[43] in order to control for individual and household level covariates, with adjustments for intra-cluster correlation due to sample design effects. Information on the covariates is collected in the survey instrument. To estimate the impact of the water and sanitation projects on the outcome variable of interest Y at the household level we will estimate

$$Y_{ijt} = \alpha_i + \lambda_t + \Delta^h_{Yt} \cdot T_{jt} + \delta(X_{ijt}) + \varepsilon_{ijt}$$

Where:
i indexes households
j indexes segments
t indexes survey wave (2011, 2012, 2013)
α_i represent a household level fixed effect (household characteristics that are time invariant)
λ_t is a vector of time period indicators, which are equal to 1 in year t
T_{jt} is the treatment indicator, equal to one for households living in segment where water and sanitation projects were implemented in the follow up and end-line surveys
Δ^h_{Yt} is the impact estimate for outcome Y (superscript *h* for household level estimate)
X_{ijt} is a vector of covariates
$\delta(.)$ is some function of the covariates
ε_{ijt} is a random disturbance with mean 0 and positive finite variance, that is allowed to be correlated across survey rounds within a segment

Two points about the estimates proposed should be emphasized: First, that the impact of the water and sanitation projects is assessed as a "compact," meaning that the treatment status is determined by the segment being located in the communities where the projects were implemented. In other words, we conduct an "intention to treat" (ITT) analysis: since take-up of water and/or sanitation services at the household level is endogenous, all households in treated segments are defined as treated, regardless of whether they take up

[43] The generalization of DID estimators are Fixed effects estimators. We discuss the fixed effects equations as fixed effect and difference in difference estimators are numerically equivalent when using 2 time periods.

services from the program. We will use "beneficiaries" to refer to any household in a treatment segment, since any are potential beneficiaries. Similarly, all households in comparison segments are classified as non-beneficiaries, even though they may have obtained piped water by other means. These results therefore provide lower bound estimates of the impact that could be realized if all complied with the treatment and there was no control group contamination. Second, not all segments originally classified as treatment segments in the 2009 initial design received projects. This implies that we may be able to improve on the matching originally performed using the initial definitions of treatment segments. We therefore will also present results using semi-parametric difference-in-differences estimators that account for the changes in treatment status in the propensity score, using the same census data, to maintain the ex-ante matching of the design.

To estimate the semi-parametric DID estimate, we need to aggregate the data to the segment level, to directly exploit the underlying assumptions of the PSM that were used to select the comparison group. At the segment level, the DID is equation is given by

$$\bar{Y}_{jt} = \alpha_j + \lambda_t + \Delta^{s}_{Yt} \cdot T_{jt} + \phi(\bar{X}_{jt}) + \varepsilon_{jt}$$

Where:
j indexes segments and t = 2011, 2012, 2013
$\bar{X}_{jt}$ is a vector of segment level covariates
T_{jt} is the treatment indicator, equal to one for segments where water and sanitation projects were implemented in the follow up and end-line surveys
α_j represent segment level fixed effects (segment characteristics that are time invariant)
λ_t is a vector of time period indicators, which are equal to 1 in year t
Δ^{s}_{Yt} is the impact estimate for outcome Y (s for segment level estimate)
$\phi(.)$ is a function
ε_{jt} is a random disturbance with mean 0 and finite positive variance, that is allowed to be correlated across survey rounds

Heckman, Ichimura, and Todd (1997) and Heckman et al. (1998) propose a DID estimator that is constructed by matching differences in pre-treatment and post-treatment outcomes for the treated to weighted averages of differences in pre-treatment and post-treatment outcomes for the comparisons. The differences are matched on the propensity score and the weights are determined non-parametrically using local linear regression or other non-parametric smoother. Using 2 time periods, baseline and a year post-intervention, the semi-parametric DID impact estimate at the segment level is given by

$$\widehat{\Delta}^{SDiD}{}_{Ytt'} = \frac{1}{\mathrm{n}_{It}} \sum_{\mathrm{i} \in \mathrm{Treat}} \left[(\bar{\mathrm{Y}}^{T}_{it} - \bar{\mathrm{Y}}^{T}_{it'}) - \sum_{\mathrm{j} \notin \mathrm{Treat}} W(P_i, P_j)(\bar{Y}^{C}_{jt} - \bar{Y}^{C}_{jt'}) \right]$$

where $t > t'$ and

$\widehat{\Delta}_{Ytt'}$ is the impact estimate of the project on outcome Y

n_{It} is the number of treatment segments in the sample at time t

$\bar{\mathrm{Y}}^{T}_{it}$ is the outcome of interest for the treatment segment i at time t

$\overline{Y}^{C}_{jt}$ is the outcome of interest for the comparison(s) segment(s) j at time t

P_i is the propensity score of segment i

$W(.)$ is a function that determines the weight of each observation in the control group.[44] In the simplest case this function selects the "nearest" neighbor, in which case the function indicate which observation j is the nearest to observation i, which is equivalent to the DID at the segment level.

To ensure that our estimates are not too sensitive to the choice of weighting function, we will present several sets of semi-parametric estimates using: kernel matching, where the counterfactual state is a weighted average of all control segments; radius-caliper matching, where the counterfactual is constructed as a weighted average of the set of control segments with propensity scores nearest to the propensity score of the treatment segment; a variation of nearest neighbor matching that avoids matches where the scores are too far apart; and, local linear matching, which is a more flexible type of kernel matching. [see Todd (2008)]

6.1.1. Separate populations and outcomes

To explore heterogeneity, for example by sex of household head and socio-economic status, we estimate the regressions interacted by these characteristics, Z_{ij}.

$$Y_{ijt} = \alpha_i + \lambda_t + \Delta_{1}{}^{h}_{Yt} \cdot T_{jt} + \Delta_{2}{}^{h}_{Yt} \cdot T_{jt} \cdot Z_{ij} + \delta(X_{ijt}) + \varepsilon_{ijt}$$

$\Delta_{2}{}^{h}_{Yt}$ gives us the additional effect for the households with (baseline) characteristics Z_{ij}.

6.1.2. Complementary Methodology

The final analysis we will explore the distinction between the intention to treat effects, i.e. effects on household within treatment segments or areas, and the average treatment on the treated, i.e. effects on the households that have a tap connection or improved latrine in treatment segment/areas. We collected information on the actual treatment status of the households (meaning if the household is connected to a water system financed by MCC WSS without conditioning on the segment). We will explore this possibility using an instrumental variable estimation, as a complementing alternative to dealing with the potential selection bias at the segment level. The instruments used will be based on the project descriptions and our discussions with stakeholders. The construction of water and sanitation schemes followed some principles which can be exploited as instruments. First, construction always began near the identified water sources. Second, segments with a higher per beneficiary municipality contribution tended to finish earlier. Third, a common reason that communities were excluded from the project was that subterranean water was found to be too contaminated to be filtered at an acceptable cost. Possible instruments are therefore distance to the water source for each household, per beneficiary municipality contribution and subterranean water contamination/quality.

[44] We note that we are not explicitly addressing the common support assumption in the discussion. The reader should assume that the *i* observations run through the common support region, and that the observations *j* run through the full set of comparisons.

7. Monitoring Plan

7.1. Adherence to treatment and control areas

The definition of control and treatment area was discussed previously; some of the segments in the sample are only partially treated and some of the comparisons segments include some treatment households. In addition, the change in the project dates was one of the main reasons that a second baseline survey needed to be scheduled, to make sure that appropriate treatment and control areas were included in the sample. The multiple survey rounds have allowed us to address deviations from the project implementation and collect information on the treatment status of the household in lieu of just the segment.

Clearly these issues have implications for the design, which we discussed when addressing the intention to treat nature of the impact evaluation design. The proposed analysis provides lower bound estimates of the impact that could be realized if all households in treatment segments were indeed treated and none of the households in the control segments were treated. To the extent possible we will explore the sensitivity of the impact estimates based on the treatment status assignment by using information on the location of the households in the sample and defining treatment areas as a function of the proximity to a WSS project (in contrast to treatment segments).

8. Administrative

8.1. Summary of Institutional Review Board requirements and clearances

Following University of Maryland Institutional Review Board Guidelines, this impact evaluation has been determined to bear minimal risk and granted exempt status. These guidelines consider minimal risk to occur when "the probability and magnitude of harm or discomfort anticipated in the proposed research are not greater, in and of themselves, than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests".

Specifically this research qualifies under the exemption[45] categories:

- *"Research involving the collection or study of existing data, documents, records, pathological or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects."*
- *"Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (a) public benefit or service programs; (b) procedures for obtaining benefits or services under those programs; (c) possible changes in or alternatives to those programs or procedures; or (d)*

[45] Exempt is defined as being exempt from further review and approval beyond the IRB Manager, or one of the two IRB Co-Chairs. It does not mean that the proposed research project was exempt from being reviewed. The research proposal was submitted to the IRB board and it was deemed to qualify for the exemption.

possible changes in methods or levels of payment for benefits or services under those programs."

8.2. Data Access, Privacy and Documentation Plan

We will produce cleaned raw datasets that follows MCC's guidelines for public use data, including programming syntax used to clean the datasets for documentation purposes.

A full set of documentation for each survey will be provided. The raw data and the data used for the final analysis will be provided. In addition, a public use version of the data files will be provided. The publicly available version will be anonymized, and thus free of personal or geographic identifiers that would permit identification of individual respondents or their household members. In addition, we will exclude variables that introduce reasonable risks of deductive disclosure of the identity of individual subjects.

In order to facilitate access to and usability of data, all datasets delivered to MCC will be accompanied by completed documentation in the form of standardized metadata using the International Household Survey Network's (IHSN) Metadata Editor.

8.3. Dissemination Plan

A report outlining the results of each survey round will be prepared; namely, a baseline report, discussing the baseline characteristics and the evaluation of the ex-ante matching strategy in February 2012; a draft interim report, estimating the short term impacts of the WSS in February 2013. Finally, we will produce an end-line report, including interim results as well, tentatively scheduled to be drafted for April 2014, which will include the analysis of the three rounds of the survey.[46]

The reports will follow a template agreed upon with MCC. The reports will validate evaluation design and revise the power analysis when necessary to verify the appropriateness of the effective sample for the impact evaluation. Estimates for figures needed in the ERR models will be included. Finally, the results will be compiled in an academic paper to be published in policy and development journals.

Presentation dissemination efforts will include: presentation of the report(s) to MCC Headquarters staff, presentation in MCC workshops, presentation of findings and key recommendations to local stakeholders, and presentation of the findings in other international development conferences.

8.4. Evaluation Team roles and responsibilities

Maureen Cropper will manage the team, which will work closely together on the WSS impact evaluation. Dr. Cropper and Dr. Raymond Guiteras will lead the technical review and be the point of contact for the evaluation, leading the analysis and reporting of findings. Miguel Almanzar will travel to the field to oversee data collection, lead the data management and econometric analysis. He will work closely with Dr. Cropper and Dr. Guiteras in the analysis, reporting, and presentation of findings.

[46] It was agreed upon in mid-2013 to include the key results from the interim report based on 2012 data in the end-line report. Only the end-line report will be finalized through MCC's formal review process and made public.

Maureen Cropper is a Distinguished University Professor of Economics at the University of Maryland, a Senior Fellow at Resources for the Future, and a former Lead Economist at the World Bank. Dr. Cropper received her Ph.D. in economics from Cornell University (1973) and has served as chair of the EPA Science Advisory Board Environmental Economics Advisory Committee and as past president of the Association of Environmental and Resource Economists. She is a member of the National Academy of Sciences and a Research Associate of the National Bureau of Economic Research. Her research has focused on valuing environmental amenities (especially environmental health effects), on the discounting of future health benefits, and on the tradeoffs implicit in environmental regulations. Her research focuses on valuing environmental programs using both stated and revealed preference approaches

Raymond Guiteras is an Assistant Professor of Economics at the University of Maryland. Dr. Guiteras received his PhD from the Massachusetts Institute of Technology in 2008 and his research focuses on issues related to the environment and health in developing countries. He is currently conducting randomized-controlled trials studying demand for latrines in rural Bangladesh, demand for and use of clean water and hand washing technologies in urban Bangladesh, and valuation of clean water in Kenya.

Miguel Almanzar is a Research Associate in the University of Maryland and is currently a Senior Research Analyst at the International Food Policy Research Institute (IFPRI), where his research focuses on the effects of public infrastructure and services on rural development. For his dissertation, Mr. Almanzar is currently analyzing the distributional impact of infrastructure provision in El Salvador. He is also evaluating the impact of nutritional and agricultural extension services in Honduras. Since 2011 he has been a part of the team working on the impact evaluation of the water and sanitation intervention in Northern El Salvador and since 2012 he has actively participated on the IFPRI team for the connectivity and electrification impact evaluations of the MCC-FOMILENIO projects. Mr. Almánzar is a PhD candidate in the Agricultural Economics and Rural Development Department of the University of Göttingen and received his MA degree in Economics from the University of Maryland-College Park. He is a native Spanish speaker and is fluent in English and French.

8.5. Budget

The total cost of the WSS activity ascends to 17.6 million. The evaluation activities amount to $295,000 in evaluators cost and $840,000 in data collection activities for a total cost of $1,135,000. The total impact evaluation costs represent 6 percent of the total cost of the water and sanitation sub-activity.

9. References

Bosch, Christophe, Kirsten Hommann, Gloria Rubio, Claudia Sadoff, and Lee Travers 2000. "Water and Sanitation." Pp. 371-404 in *A Sourcebook for Poverty Reduction Strategies*. Washington, D.C: World Bank.

Chase, Robert S. 2002. "Supporting Communities in Transition: The Impact of the Armenian Social Investment Fund." *World Bank Economic Review* 16(2):219-240.

Dehejia, R. H. and S. Wahba. 1994. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Trainning Programs." *Journal of the American Statistical Society* 1053-1062.

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parient, and Vincent Pons. 2012. "Happiness on Tap: Piped Water Adoption in Urban Morocco." *American Economic Journal: Economic Policy* 4(4):68-99.

Diggle, Peter, Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*. New York, NY: Oxford University Press.

Dirección General de Estadística y Censos. 2007. "Encuesta de Hogares de Propósitos Múltiples." Retrieved July 2013 (http://www.digestyc.gob.sv/index.php/temas/des/ehpm.html).

Donner, Allan. 1998. "Some Aspects of the Design and Analysis of Cluster Randomization Trials." *Journal of the Royal Statistical Society* 47(1).

Donner, A., N. Birkett, and C. Buck. 1981. "Randomization by cluster: sample size requirements and analysis." *American Journal of Epidemiology* (114):906-914.

Esrey, S A. 1996. "Water, waste, and well-being: a multicountry study." *American journal of epidemiology* 143(6):608-623.

Esrey, Steve A., James B. Potash, Leslie Roberts, and Clive Shiff. 1991. "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma." *Bulletin of the World Health organization* 69(5).

Fan, Victoria Y.-M. and Ajay Mahal. 2011. "What prevents child diarrhoea? The impacts of water supply, toilets, and hand-washing in rural India." *Journal of Development Effectiveness* 3(3):340-370.

Fewtrell, Lorna, Rachel B. Kaufmann, David Kay, Wayne Enanoria, Laurence Haller, and John M. Colford Jr. 2005. "Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis." *The Lancet Infectious Diseases* 5(1):42-52.

FISDL. 2005. "Fondo de Inversión Social para el Desarrollo." Retrieved 2012 (http://www.fisdl.gob.sv/documentos/libromapapobreza/).

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113(1):83-120.

Galiani, Sebastian, Martin Gonzalez-Rozada, and Ernesto Schargrodsky. 2009. "Water Expansions in Shantytowns: Health and Savings." *Economica* 76(304):607-622.

Gamper-Rabindran, Shanti, Shakeeb Khan, and Christopher Timmins. 2010. "The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach." *Journal of Development Economics* 92(2):188-200.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 1017-1098.

Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 605-654.

Heckman, J., H. Ichimura, and P. Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 261-294.

Jalan, Jyotsna and Martin Ravallion. 2003. "Does piped water reduce diarrhea for children in rural India?" *Journal of Econometrics* 112(1):153-173.

Klasen, Stephan, Tobias Lechtenfeld, Kristina Meier, and Johannes Rieckmann. 2011. "Impact Evaluation Report: Water Supply and Sanitation in Provincial Towns in Yemen." Courant Research Centre PEG, Gottingen.

Klasen, Stephan, Tobias Lechtenfeld, Kristina Meier, and Johannes Rieckmann. 2012. "Benefits trickling away: the health impact of extending access to piped water and sanitation in urban Yemen." *The Journal of Development Effectiveness* 4(4):537-565.

Koolwal, Gayatri and Dominique van de Walle. 2013. "Access to Water, Women's Work, and Child Outcomes." *Economic Development and Cultural Change* 61(2):369 - 405.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix P. Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *The Quarterly Journal of Economics* 126(1):145-205.

Lamichhane, Dirga K. and Eiji Mangyo. 2011. "Water accessibility and child health: Use of the leave-out strategy of instruments." *Journal of Health Economics* 30(5):1000-1010.

Lokshin, M. and R. Yemtsov. 2005. "Has rural infrastructure rehabilitation in Georgia helped the poor?" *The World Bank Economic Review* 19(2):311-333.

Mangyo, Eiji. 2008. "The effect of water accessibility on child health in China." *Journal of Health Economics* 27(5):1343-1356.

MCC. 2012. "Millennium Challenge Account – El Salvador, Monitoring and Evaluation Plan." Version 6. (https://www.mcc.gov/documents/data/SLV_ME_Plan_Sept_2012_MCC_Approved.pdf).

Miguel, Edward and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1):159-217.

Murray, D. 1998. "Sample Size Detectable Difference, and Power." Pp. 349-413 in *Design and Analysis of Group-Randomized Trials*. New York and Oxford: Oxford University Press.

Nauges, Celine and Jon Strand. 2007. "Estimation of non-tap water demand in Central American cities." *Resource and Energy Economics* 29(3):165-182.

Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose L. Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16(2):241-274.

Olmstead, Sheila M., Michael W. Hanemann, and Robert N. Stavins. 2007. "Water demand under alternative price structures." *Journal of Environmental Economics and Management* 54(2):181-198.

Pattanayak, Subhrendu K., Christine Poulos, Jui-Chen Yang, and Sumeet Patil. 2010. "How valuable are environmental health interventions? Evaluation of water and sanitation programmes in India." *Bulletin of the World Health Organization* (88):535-542.

Pattanayak, SK, C Poulos, J-C Yang, SR Patil, and KJ Wendland. 2009. "Of taps and toilets: quasi-experimental protocol for evaluating community-demand-driven projects." *Journal of Water and Health* 7(3):434-451.

Pattanayak, S.K., J.-C. Yang, C. Poulos, S.R. Patil, B. Arnold, and J.M. Colford. 2008. "Three years later: Environmental health impacts of a community-demand-driven water and sanitation program in rural Maharashtra, India." *RTI International Working Paper 08-01*.

Pattanayak, Subhrendu K., Jui-Chen Yang, Dale Whittington, and K. C. Bal Kumar. 2005. "Coping with unreliable public water supplies: Averting expenditures by households in Kathmandu, Nepal." *Water Resources Research* 41(2).

Poulos, C., S.K. Pattanayak, and K. Jones. 2006. "A Guide to Water and Sanitation Sector Impact Evaluations." *Doing Impact Evaluation Series, World Bank*.

Purdon, Susan. 2002. "Estimating the impact of labour market programmes." Great Britain.

Rauniyar, Ganesh, Aniceto Orbeta, and Guntur Sugiyarto. 2011. "Impact of water supply and sanitation assistance on human welfare in rural Pakistan." *Journal of Development Effectiveness* 3(1):62-102.

Ravallion, M. 2008. "Evaluating Anti-Poverty Programs." in *Handbook of Development Economics*, edited by T. Paul Schultz and John Strauss. Amsterdam.

Rosenbaum, P.R. and D. B. Rubin. 1983. "The Central Role of Propensity Score in Observational Studies." *Biometrika* 41-55.

Rubin, D. 1974. "Estimating Causal Effects to Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688-701.

Rubin, D. 1977. “Assignment to Treatment Group on the Basis of a Covariate.” *Journal of Educational Studies.* 2:1-26.

Rubin, D. 1979. “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies.” *Journal of the American Statistical Association* 74:318-328.

Strand, Jon and Ian Walker. 2005. “Water markets and demand in Central American cities.” *Environment and Development Economics* 10(03):313-335.

Todd, Petra E. 2008. “Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated.” in *Handbook of Development Economics*, edited by T. Paul Schultz & John A. Strauss. 1st ed.

WHO. 2013. “media Center: Fact sheets: Diarrhoeal disease.” Retrieved July 2013 (http://www.who.int/mediacentre/factsheets/fs330/en/).

WHO/UNICEF. 2005. “Water for life : making it happen.”. (http://www.who.int/water_sanitation_health/waterforlife.pdf).

WHO; UNICEF. 2013. “Progress on sanitation and drinking-water. 2013 update: Joint Monitoring Programme for Water Supply and Sanitation.”. Retrieved August 2013 (http://www.who.int/water_sanitation_health/publications/2013/jmp_report/en/).

Annex 1: List of eligible municipalities

DEPARTMENT	MUNICIPALITY
CABAÑAS	DOLORES
CABAÑAS	GUACOTECTI
CABAÑAS	ILOBASCO
CABAÑAS	SAN ISIDRO
CABAÑAS	SENSUNTEPEQUE
CABAÑAS	TEJUTEPEQUE
CABAÑAS	VICTORIA
CHALATENANGO	AGUA CALIENTE
CHALATENANGO	AZACUALPA
CHALATENANGO	CITALA
CHALATENANGO	COMALAPA
CHALATENANGO	CONCEPCION QUEZALTEPEQUE
CHALATENANGO	DULCE NOMBRE DE MARIA
CHALATENANGO	EL CARRIZAL
CHALATENANGO	EL PARAISO
CHALATENANGO	LA PALMA
CHALATENANGO	LA REINA
CHALATENANGO	LAS FLORES (San José)
CHALATENANGO	NOMBRE DE JESUS
CHALATENANGO	NUEVA TRINIDAD
CHALATENANGO	SAN ANTONIO LA CRUZ
CHALATENANGO	SAN FRANCISCO LEMPA
CHALATENANGO	SAN IGNACIO
CHALATENANGO	SAN LUIS DEL CARMEN
CHALATENANGO	SAN MIGUEL DE MERCEDES
CHALATENANGO	SAN RAFAEL
CHALATENANGO	SANTA RITA
CHALATENANGO	TEJUTLA
CHALTENANGO	NUEVA CONCEPCION
CUSCATLAN	SUCHITOTO
LA LIBERTAD	SAN PABLO TACACHICO
LA UNION	ANAMOROS
LA UNION	BOLIVAR
LA UNION	EL SAUCE
LA UNION	LISLIQUE
LA UNION	NUEVA ESPARTA
LA UNION	POLOROS
LA UNION	SAN JOSE (de la fuente)
MORAZAN	ARAMBALA
MORAZAN	CACAOPERA
MORAZAN	CHILANGA

MORAZAN	CORINTO
MORAZAN	DELICIAS DE CONCEPCION
MORAZAN	EL DIVISADERO
MORAZAN	EL ROSARIO
MORAZAN	JOCOAITIQUE
MORAZAN	LOLOTIQUILLO
MORAZAN	MEANGUERA
MORAZAN	OSICALA
MORAZAN	PERQUIN
MORAZAN	SAN FERNANDO
MORAZAN	SENSEMBRA
MORAZAN	SOCIEDAD
MORAZAN	YAMABAL
SAN MIGUEL	CIUDAD BARRIOS
SAN MIGUEL	NUEVO EDEN DE SAN JUAN
SAN MIGUEL	SAN GERARDO
SAN MIGUEL	SAN LUIS DE LA REINA
SAN MIGUEL	SESORI
SAN SALVADOR	EL PAISNAL
SANTA ANA	SAN ANTONIO PAJONAL
SANTA ANA	SANTA ROSA GUACHIPILÍN

Annex 2: Project Selection Criteria

TABLE 15 CRITERIA FOR SELECTION OF PROJECT PROFILES FOR THE FEASIBILITY STAGE

CRITERIA		
1	**ELIGIBLE MUNICIPALITIES**	According to Integrated Municipal Marginality Index the municipalities to participate are the High and Extreme Poverty Moderate Extreme Poverty
2	**COMMUNITY CONTRIBUTION**	Municipal Agreement which ensures the contribution of 10% cash of the total project investment, by the municipality General Assembly Act which ensures the contribution of 10% in cash or in-kind of the total project investment, by the community
3	**SOCIAL FEASIBILITY**	COMMUNITY ORGANIZED: Legal documentation of organization or Act General Assembly showing willingness to organize. Community commitment to participate in all the steps of the project, especially in decision-making activities. (Minutes of General Assemblies) Community-municipal commitment to the management, operation and maintenance of the works and system. The communication mechanism will be through General Assemblies which shall record the agreements reached, formalizing these.
4	**COST PER BENEFICIARY**[47]	The ratio of the cost of the project and the number of beneficiaries is maximum $ 850.00 per beneficiary

[47] Beneficiary: People with access to potable water and / or sanitation as result of the intervention of the MCC compact.

TABLE 16 CONDITIONS PROJECTS MUST MEET IN THE FEASIBILITY PHASE

	CRITERIA	
1	**ELIGIBLE PROJECTS**	All those selected profiles with the above criteria.
2	**ECONOMIC AND FINANCIAL VIABILITY**	Provision of sustainable monthly fee paid to be established in set community-municipality-contractor Feasibility Study. The rates must cover the costs of operating and maintaining the system Each individual project shall not exceed $ 500,000. From higher cost proposed works will be considered on a case by case basis by the Technical Committee comprising representatives of FISDL and FOMILENIO, taking as an observer to the MCC. FOMILENIO The Board will approve the decision based on the expected impact on the population to benefit.
3	**TECHNICAL FEASIBILITY**	Provision of Design: 70-100 liters / person / day for distribution household. Provision of Design: Minimum of 35 liters / person / day for public tap (cantarera) distribution. Water Quality: Drinking Availability of Source Water for a period of 20 years VIP latrines will be constructed (Improved Pit Latrines) when ground conditions permit (not rocky terrain, water table, etc..). If these conditions are not met, will provide composting latrines type. They must respect the standards set by the Ministry of Health. In addition shall comply with the contents of the Folder Guide Formulation Techniques for Drinking Water Projects FISDL, see extract requirements FISDL Formulation Guide Annex II.
4	**LEGAL FEASIBILITY**	Ownership of land in favor of the municipalities or engagement letter purchase / donation / loan of land where the water source and where it will build the works (with no legal or social issues) and authorization of easements.
5	**ENVIRONMENTAL FEASIBILITY**	Water Source: Able to supply the amount of 70 to 100 liters / person / day for the design period (20 years) MARN Permissions: Environmental Impact Data necessary Impact environmental Micro-Management Plan

Annex 3: List and description of projects

Deleted to protect the anonymity of subjects in the evaluation

Annex 4: Matching Results

TABLE 17 MATCHING EQUATION AND RESULTS

Logistic regression for propensity score	88 Treatment Segments-2009	65 Treatment Segments-2011
	Initial	***Updated***
Municipal Pop. 2005	-0.0076	-0.007
	[0.0017]***	[0.0018]***
Density	-0.0015	-0.0013
	[0.00066]*	[0.00061]*
Surface area occupied by water	0.59	0.69
	[0.37]	[0.37]
Avg. Temperature	-0.12	-0.12
	[0.062]*	[0.064]
Household Size	-1.07	-0.96
	[0.30]***	[0.30]**
Unimproved Water	1.64	1.37
	[0.72]*	[0.74]
Private Well	2.58	2.44
	[0.70]***	[0.70]***
Piped Water	1.07	0.94
	[0.51]*	[0.51]
No Sanitation	0.26	-0.066
	[0.93]	[0.97]
Composting Latrine	1.93	1.88
	[0.63]**	[0.63]**
% in-Home Business	-3.06	-3.24
	[1.55]*	[1.63]*
Number Emigrated	0.24	-0.044
	[0.43]	[0.45]
Dep_1	-1.13	-0.62
	[0.87]	[0.88]
Dep_2	-0.49	-0.26
	[0.38]	[0.40]
Dep_7	-1.97	-2.3
	[0.82]*	[1.09]*
Dep_8	-1.06	-0.71
	[0.43]*	[0.45]
Dep_9	-1.43	-0.9

	[0.55]**	[0.55]
Constant	5.95	5.25
	[2.38]*	[2.44]*
Observations	1047	1047
*** p<0.1 ** p<0.05 ***p<0.01**		

TABLE 18 MATCHED AND UNMATCHED SAMPLES – BALANCING TESTS AT 2009 BASELINE

		Mean		Bias		t-test		
Variable		**Treated**	**Control**	**%bias**	**%reduct**	**t**	**p>t**	
Departments	Unmatched	0.02273	0.01877	2.8		0.26	0.795	
1	Matched	0.02273	0.01136	7.9	-187.1	0.58	0.563	
	Unmatched	0.45455	0.25443	42.6		4.07	0	***
2	Matched	0.45455	0.52273	-14.5	65.9	-0.9	0.368	
	Unmatched	0.02273	0.09385	-30.7		-2.26	0.024	**
7	Matched	0.02273	0.01136	4.9	84	0.58	0.563	
	Unmatched	0.20455	0.19812	1.6		0.14	0.885	
8	Matched	0.20455	0.20455	0	100	0	1	
	Unmatched	0.10227	0.13347	-9.7		-0.83	0.407	
9	Matched	0.10227	0.07955	7	27.2	0.52	0.602	
% Farmer	Unmatched	0.4109	0.36915	16.1		1.36	0.174	
	Matched	0.4109	0.41858	-3	81.6	-0.21	0.836	
% Agriculture	Unmatched	0.47088	0.42101	18		1.53	0.127	
	Matched	0.47088	0.46947	0.5	97.2	0.04	0.971	
% Animal Husbandry	Unmatched	0.20076	0.18993	4.7		0.43	0.67	
	Matched	0.20076	0.19259	3.6	24.6	0.23	0.816	
% in-Home Business	Unmatched	0.07284	0.09671	-27		-2.02	0.043	**
	Matched	0.07284	0.0766	-4.3	84.3	-0.34	0.735	
Municipal Pop. 2005	Unmatched	9874.1	20797	-76.8		-5.32	0	***
	Matched	9874.1	10180	-2.2	97.2	-0.28	0.782	
% with Electricity	Unmatched	0.7807	0.78211	-0.7		-0.06	0.953	
	Matched	0.7807	0.76403	8.5	-1083.8	0.53	0.593	
Density	Unmatched	85.725	318.99	-50.4		-3.42	0.001	***
	Matched	85.725	90.799	-1.1	97.8	-0.19	0.851	

Variable		Mean		Bias		t-test		
		Treated	Control	%bias	%reduct	t	p>t	
Area	Unmatched	5.4309	3.6605	36.9		4.08	0	***
	Matched	5.4309	4.9322	10.4	71.8	0.67	0.504	
Avg. Temperature	Unmatched	24.782	24.702	3.6		0.37	0.714	
	Matched	24.782	25.057	-12.5	-243.8	-0.77	0.441	
Avg. Precipitation	Unmatched	1875.4	1910	-18.4		-1.84	0.066	*
	Matched	1875.4	1871.8	1.9	89.4	0.12	0.901	
Household Size	Unmatched	4.322	4.5062	-32.8		-2.84	0.005	***
	Matched	4.322	4.3679	-8.2	75.1	-0.58	0.566	
No Sanitation	Unmatched	0.19452	0.1976	-1.7		-0.14	0.888	
	Matched	0.19452	0.18962	2.8	-59	0.19	0.848	
Unimproved Water	Unmatched	0.16223	0.1335	13.3		1.21	0.228	
	Matched	0.16223	0.15375	3.9	70.5	0.23	0.815	
Private Well	Unmatched	0.2133	0.11625	48.3		4.66	0	***
	Matched	0.2133	0.22093	-3.8	92.1	-0.2	0.842	
Piped Water	Unmatched	0.53625	0.66159	-39.9		-3.5	0	***
	Matched	0.53625	0.55294	-5.3	86.7	-0.33	0.738	
Private Piped water	Unmatched	0.48022	0.59863	-38.1		-3.27	0.001	***
	Matched	0.48022	0.49301	-4.1	89.2	-0.27	0.789	
Private piped<Daily	Unmatched	0.34404	0.3289	4.8		0.42	0.678	
	Matched	0.34404	0.32988	4.5	6.5	0.29	0.774	
Latrine normal	Unmatched	0.34936	0.39067	-17.1		-1.57	0.118	
	Matched	0.34936	0.33741	4.9	71.1	0.32	0.751	
Composting Latrine	Unmatched	0.27264	0.13817	56.3		5.51	0	***
	Matched	0.27264	0.29825	-10.7	81	-0.63	0.531	
Sewer	Unmatched	0.02526	0.09872	-43.2		-3.02	0.003	***
	Matched	0.02526	0.03843	-7.7	82.1	-0.84	0.4	
Dirt Floors	Unmatched	0.37558	0.36117	6.8		0.56	0.575	
	Matched	0.37558	0.37528	0.1	97.9	0.01	0.992	

*** p<0.1 ** p<0.05 ***p<0.01**

TABLE 19 BIAS DISTRIBUTION STATISTICS - TESTS AT 2009 BASELINE

Sample	Pseudo R2	LR chi2	p>chi2	Mean	Med

				Bias	Bias
Raw	0.216	130.77	0	23.9	18
Matched	0.053	13.05	0.984	5.2	4.3

TABLE 20 MATCHED AND UNMATCHED SAMPLES – BALANCING TESTS AT 2011 BASELINE

		Mean		Bias		t-test		
Variable		**Treated**	**Control**	**%bias**	**%reduct**	**t**	**p>t**	
Departments	Unmatched	0.02	0.02	3.80		0.35	0.72	
1	Matched	0.02	0.05	-16.40	-337.30	-0.83	0.41	
	Unmatched	0.46	0.29	36.20		3.37	0.00	***
2	Matched	0.46	0.53	-14.80	59.20	-0.92	0.36	
	Unmatched	0.01	0.09	-36.10		-2.50	0.01	**
7	Matched	0.01	0.00	5.40	84.90	1.00	0.32	
	Unmatched	0.21	0.18	6.90		0.63	0.53	
8	Matched	0.21	0.21	0.00	100.00	0.00	1.00	
	Unmatched	0.12	0.13	-5.00		-0.44	0.66	
9	Matched	0.12	0.11	3.50	29.40	0.24	0.81	
% Farmer	Unmatched	0.41	0.38	10.90		0.91	0.36	
	Matched	0.41	0.42	-4.30	60.90	-0.27	0.79	
% Agriculture	Unmatched	0.47	0.43	14.10		1.18	0.24	
	Matched	0.47	0.47	-0.30	98.20	-0.02	0.99	
% Animal Husbandry	Unmatched	0.21	0.20	5.00		0.45	0.65	
	Matched	0.21	0.22	-2.20	56.80	-0.14	0.89	
% in-Home Business	Unmatched	0.07	0.09	-26.10		-1.93	0.05	*
	Matched	0.07	0.08	-8.30	68.20	-0.62	0.54	
Municipal Pop. 2005	Unmatched	10065.0	19517.0	-68.30		-4.67	0.00	***
	Matched	10065.0	9298.8	5.50	91.90	0.75	0.46	
% with Electricity	Unmatched	0.79	0.77	7.40		0.59	0.55	
	Matched	0.79	0.73	27.20	-269.40	1.60	0.11	
Density	Unmatched	90.93	285.48	-44.20		-2.96	0.00	***
	Matched	90.93	85.89	1.10	97.40	0.18	0.86	
Area	Unmatched	5.41	3.82	32.80		3.53	0.00	***
	Matched	5.41	5.53	-2.50	92.40	-0.15	0.88	

		Mean		Bias		t-test		
Variable		Treated	Control	%bias	%reduct	t	p>t	
Avg. Temperature	Unmatched	24.88	24.68	9.10		0.87	0.38	
	Matched	24.88	24.32	25.30	-179.70	1.39	0.17	
Avg. Precipitation	Unmatched	1870.1	1912.4	-22.40		-2.20	0.03	**
	Matched	1870.1	1903.2	-17.50	21.80	-1.01	0.31	
Household Size	Unmatched	4.32	4.51	-33.10		-2.81	0.01	***
	Matched	4.32	4.37	-8.50	74.50	-0.57	0.57	
No Sanitation	Unmatched	0.19	0.21	-11.40		-0.90	0.37	
	Matched	0.19	0.19	-0.30	97.40	-0.02	0.98	
Unimproved Water	Unmatched	0.15	0.14	7.20		0.64	0.52	
	Matched	0.15	0.17	-5.90	18.70	-0.36	0.72	
Private Well	Unmatched	0.22	0.12	49.00		4.85	0.00	***
	Matched	0.22	0.18	17.90	63.40	1.04	0.30	
Piped Water	Unmatched	0.55	0.65	-34.60		-2.99	0.00	***
	Matched	0.55	0.57	-7.20	79.10	-0.46	0.65	
Private Piped water	Unmatched	0.48	0.59	-36.10		-3.04	0.00	***
	Matched	0.48	0.52	-12.40	65.60	-0.78	0.43	
Private piped<Daily	Unmatched	0.34	0.33	2.60		0.22	0.83	
	Matched	0.34	0.35	-5.20	-100.00	-0.33	0.74	
Latrine normal	Unmatched	0.35	0.39	-16.00		-1.46	0.14	
	Matched	0.35	0.36	-6.90	57.20	-0.45	0.66	
Composting Latrine	Unmatched	0.29	0.14	57.50		5.71	0.00	***
	Matched	0.29	0.27	8.70	84.90	0.52	0.60	
Sewer	Unmatched	0.02	0.09	-40.30		-2.79	0.01	***
	Matched	0.02	0.04	-8.00	80.20	-0.87	0.38	
Dirt Floors	Unmatched	0.36	0.37	-5.50		-0.44	0.66	
	Matched	0.36	0.42	-25.90	-371.10	-1.64	0.10	

* p<0.1 ** p<0.05 ***p<0.01

TABLE 21 BIAS DISTRIBUTION STATISTICS - TESTS AT 2011 BASELINE

Sample	KS test Equality Distribution of Scores	Pseudo R2	LR chi2	p>chi2	Mean Bias	Med Bias

Raw	0.000	0.18	111.89	0.00	23.40	16.00
Matched	0.920	0.08	18.06	0.84	8.90	6.90

Annex 5: Water quality testing

In all households interviewed for this evaluation, we will conduct a field test of residual chlorine levels in the household's drinking water supply. This test alone, however, is not enough to show whether or not water poses a health risk to the households. Non-chlorinated water may have no bacterial contamination, for example. To have better insight into the health risk posed by the water supply before and after the Compact investments, additional laboratory-based tests of water quality will be conducted in a sample of households. Also, the water quality in piped systems and some other non-piped sources will be tested in a subsample of communities.

These additional tests will be done in two ways. First, the Ministry of Health (through personnel of the health posts) conducts tests for bacteriological contamination in piped water systems. We expect to compile these results during the end-line survey, with the health post survey.

Second, a private firm will test drinking water for bacteriological contamination. The tests of household drinking water will be done on stored water, if the household stores drinking water, or will be taken directly from the tap or pipe if the household does not store drinking water. For the tests of source water quality, the firm will search for the sources used by the households included in the water quality testing program.

Testing will be done in April – May each year of the survey, as close as possible to the timing of the household survey. The lab technicians will fill out a small survey at each household visited to assist in the process of linking household survey results and the results of household-level and source-level water quality tests.

The sample size of water quality tests will not necessarily give us sufficient power to draw conclusions about water quality across the whole sample. The idea is to get a sense of how important water quality problems are in the project and control areas and whether problems lie in source contamination, in in-house contamination, or both.

<u>Water Testing Procedure</u>

They will then follow the following procedure to identify the number of sources to test and which sources to test:

Drinking water quality tests among households are interviewed in 66 census segments and 12 households in each segment. These tests will include a sample from the home and the source of the water supply (see Table 22). The samples will be taken from either the home or the source to a laboratory to be analyzed.

TABLE 22 SAMPLE SIZE FOR WATER QUALITY TESTS

Number of Clusters	**66 census segments**
Sample from household drinking water	66 x 12 = 792
Sample from source **(between 3 and 6 per census segment)**	Between 198 - 396

All of the tests will be a microbiological analysis of the water as well as a residual chlorine analysis using a colorimeter. The scope of the microbiological analysis of the water samples should include: total number of coliform bacteria and fecal coliforms and the presence of E. Coli.

i) **The laboratory personnel will take 12 drinking water samples** from the a list of homes in each census segment, according to the following:

- If the home uses some device for storing water (canteen, bucket) take a sample from this storage device.
- If the home obtains its water **directly** from its own or a neighbor's tap, a private well, a hose or pipeline from a spring, take the simple directly from the source. In the case of pipelines or hoses it is not necessary to conduct the test in the initial source, except from a tap or hose. For artificial wells and other water sources (watering hole, spring, stream, river) go to the source for the sample.

ii) **The laboratory personnel will take a minimum of 3 and a maximum of 6 water samples** at the sources these households use, according to the following:

In segments where the principal source of water is a piped water system, take 3 samples from household tap. Of the 12 homes choose 3: the closest to the storage tank, one mid-way and the furthest.

- Take the 3 remaining water samples from other water sources that do not have a connection to their home. In the case that there is no other non-piped water source in a segment, only 3 water samples in this segment are taken.
- *If there are public taps in the segments samples are always taken from these sources*

In segments without piped water systems, take up to 6 water samples from the household's sources. If there are not six different sources, take one sample from each source.

- If the 12 households get water from more than 6 different sources, take a sample from the 6 most commonly used sources.
- Take one sample per source (even if 2 or more households use the same source). Take a maximum of 6 samples in these segments.

iii) Fill a short survey to identify the home, the corresponding geocode and the sample associated with the home (for both drinking water and the source).

iv) Use the equipment to transport the water samples for the microbiological analysis and residual chlorine, in the time required by the laboratory so that the analysis will be valid.

LOCATIONS OF WATER QUALITY TESTING

Deleted to protect the anonymity of subjects in the evaluation