MCC – Lesotho
Rural Water and Sanitation Project

Evaluation Design and Data Collection Needs

Revised: February 23, 2015

NORC at the UNIVERSITY of CHICAGO

Table of Contents

Abbreviations

AMP	Activity Monitoring Plan (questionnaire)
BOS	Lesotho Bureau of Statistics
CLO	Community Liaison Officer
CTV	Continuous treatment variable
DD	Difference in differences
DHS	Department of Health Services
DRWS	Department of Rural Water and Sanitation
DQA	Data Quality Assessment
EA	Census enumeration area
HALEH	Household Awareness Latrine and Environmental Hygiene
HH	Household
IFB	Information for Bid
ITT	MCC Indicator Tracking Table
MCA	Millennium Challenge Account
M&E	Monitoring and Evaluation
MDES	Minimum detectable effect size
NORC	NORC at the University of Chicago
O&M	Operations and maintenance
PDNA	Engineering company responsible for Urban Water oversight and reporting
PHAST	Participatory Health and Sanitation Training
PSU	Primary statistical unit
RCT	Randomized control trial
RDD	Regression discontinuity analysis
RSA	Researcher Supplemental Application
TBD	To be determined
TOR	Terms of reference
ToT	Treatment of the treated
VIP	Ventilated Improved Pit
VWHC	Village Water and Health Committee
WASCO	Water and Sanitation Company
WHO	World Health Organization
WRD	Water-Related Disease

Page intentionally left blank

1. INTRODUCTION

Lesotho is making significant strides in providing improved water and sanitation for its population. The direct benefits from improved water supply are well-researched, and should directly reduce the burden of disease and time spent caring for the sick. Moreover, it is estimated that in Africa people spend 40 billion hours every year collecting water.[1] Improving access to safe drinking water and basic sanitation would, therefore, lead to wide-ranging social and economic benefits to Lesotho's citizens, providing substantial time savings and widening opportunities for income generation.

Through its Compact with the Government of Lesotho the Millennium Challenge Corporation (MCC) awarded $164-million over five years for investment in improved water supplies and sanitation facilities for rural and urban domestic, commercial, and industrial users.

This report focuses on MCC's rural water interventions, which comprise of interventions designed to improve the rural water systems with the objective of providing safe drinking water close to households. It is anticipated that the intermediate outcomes of improved water sources will both decrease the burden of water-borne illness by creating a cleaner more sanitary water supply and also significantly cut down on the amount of time it takes households to collect water. This, in turn, should help the MCC fulfill its mandate of reducing poverty through economic growth by allowing household beneficiaries to be more productive and generate more income than they would have in the absence of improved access to cleaner water.

As part of its commitment to transparently and thoroughly monitor and evaluate its activities, the MCC contracted NORC in 2007 to conduct an impact evaluation of its water sector activities. To date, NORC has submitted to MCC two Evaluation Design Reports for the water sector activities; the most recent of these, in June 2012. Since then, however, several new developments in the implementation of the rural water activities have occurred. First, significant implementation delays occurred in the construction of new and improved water systems in rural villages. However, project implementation is now complete, and it is possible to assess the implications of these delays on the existing evaluation design. Second, MCC has recently (in 2013) constructed a Program Logic diagram for the rural water component of the MCA Lesotho Compact (see Annex A). As such, this reports updates, as needed, and links the June 2012 evaluation design to the new program logic diagram.

In aiming to reach the Millennium Development Goals by 2015, Lesotho faces a particular challenge in improved rural water delivery which has remained relatively static at 75-77 percent in the 22 years between 1990 and 2012. Over the same period there has been an even greater challenge in improved sanitation even though coverage grew from 20 to 27 percent.[2] The national figures on improved sanitation coverage in Africa have been described as "grim" by international

[1] www.charitywater.org/whywater/

[2] Estimates on the use of Water Sources and Sanitation Facilities, updated April 2014. WHO/UNICEF Joint Monitoring Program for Water Supply and Sanitation.
http://www.wssinfo.org/documents/?tx_displaycontroller%5Btype%5D=country_files

bodies.[3] In light of an urban sector where 93 percent of households have improved water sources, a major thrust for improved services is clearly needed in the rural sector.

The marked discrepancy in the figures between water and sanitation coverage also points to the additional challenge largely, but not exclusively, in the rural sector. Attention tends to focus on delivery of water systems without simultaneous attention being paid to hygiene promotion and sanitation facilities. A strategy to reduce water-related disease (WRD) through water and sanitation has to ensure that the three components: hygiene promotion, sanitation facilities and water systems have to be combined to achieve the desired impact.[4]

2. OVERVIEW OF THE COMPACT AND THE INTERVENTIONS EVALUATED

2.1 Overview of the Project Implementation

2.1.1 Program Description

The Lesotho Rural Water and Sanitation Project (WSP) in rural areas provided for improved water and sanitation services for 27,245 households or approximately 160,000 persons through construction of new water systems and ventilated improved pit (VIP) latrines. These households are located in 250 villages that were identified by the Department of Rural Water Supply (DRWS) as lacking access to safe drinking water and adequate sanitation. Much of the existing rural water supply system in Lesotho is, for a variety of reasons, poorly maintained and the small scale water systems often do not work continuously.[5] Within this backdrop, the MCC intervention sought to build water systems, while relying on DRWS efforts to ensure their maintenance and sustainability through training Village Water and Health Committees and Water Minders in a range of management and maintenance skills.

In addition to MCC-funded construction of new water systems and VIP latrines, DRWS also provided Participatory Hygiene Awareness and Sanitation Training (PHAST)[6] and Aftercare training to participating villages. PHAST, which occurred before the construction of water systems commenced, consisted of two components: training for the entire community and a training workshop for Village Water and Health Committees (VWHC). PHAST training was the responsibility of DRWS and was provided to communities by DRWS District Community Liaison Officers (CLO).

[3] www.unicef.org/wash/files/gafull.pdf

[4] Cairncross, Sandy, et al. Water, sanitation and hygiene for the prevention of diarrhea. International Journal of Epidemiology 2010;39:i193–i205.

[5] Lesotho Country Proposal to the Millennium Challenge Corporation (MCC). July 2006. A Programme for Improvement of Water Supply, Rehabilitation of Health Infrastructure and Promotion of Private Business Development, see p26 which mentions aging capacity and p50 which mentions gravity fed systems.

[6] PHAST and Aftercare trainings in Phase-A villages and PHAST in Phase-C villages were jointly funded by DRWS and MCA. MCA provided funds to DRWS to provide snacks to community members and lunches for the VHWC during PHAST, as well as per diems for the CLO to cover the cost of paying a village household for lodging. DRWS paid the CLO their salary and used government resources (car and petrol) to get to and from the village.

Community-wide hygiene awareness and sanitation training: Delivered to the entire community by a CLO, this training consists of a participatory approach in which the CLO conducts a transact walk through the village with the entire community. During this walk, the CLO raises awareness about hygiene and sanitation by pointing out examples unhygienic/unsanitary practices and informing them about solutions they must implement to change those practices and improve hygiene within the community.

Training to VWHCs: After the community-wide PHAST is completed, the community democratically elects the VWHC. The role of the VWHC, of which the Water Minder[7] is a member, is to serve as a source of information to the community on benefits of access to clean water, water-related disease control, disposal of dirty water/waste water management, types of latrines and their requirements, among other hygiene and sanitation topics. CLOs do not use a formal manual, but instead, use a series of pictures (about 80 pictures with Sesotho script) to train the VWHC on good hygiene practices. The VWHC members are also tasked with helping community members build their hand-washing models (tippy-taps, for example) and soak away pits, and informing them of preparations required to receive a VIP latrine. Towards this end, the DRWS CLOs train VWHC members on positive hygiene and sanitation practices and teach them how to build hand-washing models/tippy taps. The VWHC treasurer also received training on keeping an account book.

As mentioned before, PHAST generally took place in the pre-construction phase, and served as one indicator of a village's "readiness" for construction of a water system. In most of the villages, VWHC training also took place before construction began. There are a few exceptions: one village each in Phase A and Phase C received PHAST community training after construction began, while two village in Phase A and 18 villages in Phase C received VWHC training after the start of construction. However, for the most part, these trainings occurred in the pre-construction stages.

In addition to PHAST training, DRWS was also responsible for providing Aftercare Training to VWHCs. In keeping with the World Bank strategy toward water supply, the DRWS Aftercare Strategy aims to put in place institutional and financial mechanisms to sustain the construction of water supplies for their 10-15 year design life.

Aftercare Training occurs after the construction of the village Water & Sanitation System. As described in the DRWS Community Management Handbook, the Aftercare Training is intended to build the VWHC capacity to perform all operation and maintenance activities on the village water and hygiene system.

Separate from the DRWS Aftercare Training, all village Water-Minders were also supposed to receive on-site training from the building contractor to learn to operate their village water system. This training should have occurred during the construction of the water system. Water Minders were expected to participate in the construction process, so that they are well informed about the make-up of the system. After the completion of the system, the Water Minder was

[7] The Village Water Minder is a member of the VWHC. His/her primary responsibility within the committee is to identify and report maintenance problems that require attention to the VWHC. The Water Minder also presents to the committee a cost estimate for fixing the problem at hand, so that the VWHC treasurer can provide her/him with the required funding to buy parts and repair the system.

supposed to receive a copy of the Operation & Maintenance Manual for the water system written by the contractor, as well as a toolkit for maintenance functions.[8]

2.1.3 Project Implementation

The Lesotho Rural WSP was implemented by the Department of Rural Water Supply (DRWS) in all 10 districts of Lesotho. In each district, DRWS identified and selected 25 villages that lacked access to safe drinking water and adequate sanitation for construction of new rural water and hygiene systems and accompanying hygiene training activities. In August 2008, NORC facilitated the random assignment of 100 of the 250 villages (10 per district) into treatment and control villages for the impact evaluation. The 100 villages were selected by representatives from the 10 districts on the basis of their "readiness" to receive construction. For a village to be considered "ready" (1) it must have undergone PHAST training and have a fully functioning VWHC committee and (2) a feasibility study must have been undertaken by a DRWS engineers to determine the type of water system needed for that particular village, the design scheme, and the estimated project cost.[9]

The selection process of the 100 study villages has implications for the external validity of the evaluation. In particular, one would want to know whether the 50 villages assigned to treatment were not just representative of those in the control group but also of those in the 150 remaining villages selected for later assistance. Currently, we do not know the extent to which these 150 villages are different or similar to the Phase-A villages in a systematic way that would be correlated with the former's response to treatment. Since the 100 villages assigned to treatment and control were selected on the basis of a set of "readiness" to receive construction criteria, it is possible that they were more advanced, had more motivated leadership and/or communities, or were different to the non-selected 150 villages in a systematic way. This could be verified through a two-pronged strategy, if it is of interest to MCC. Our proposed approach to determine the external validity of the evaluation design would involve first using existing datasets (IEMS, CMS, DHS, census, etc.) to explore conditions/characteristics in as many of the Phase-B villages and villages that are not within the purview of MCC's compact. Second, we would conduct further web and donor-document research as well as correspond by phone and email with key informants to explore conditions/characteristics the degree to Phase-B and Phase-A villages were different. The expectation would be to apply suitable parametric adjustments to the results of the Phase-A-and-C-based present design so as to extend their external validity to villages in Phase B.

The random assignment of the 100 villages to treatment and control status was conducted in a public event to assure transparency. Fifty villages (5 in each district) were randomly selected for the first wave of project implementation (Phase A); this group constituted the treatment group.[10] The remaining 50 villages from the 10 districts were assigned to the control group (Phase C).

[8] Aftercare has been completed for all Phase-A villages; however, we do not have dates of completion. Further, we have spoken with representatives from Cowater and DRWS, and neither knows whether Aftercare Training was completed for Phase-A_1 and Phase-C villages. However, we will gather this information from the VWHC surveys, if they are conducted. Similarly, the endline Water Minder questionnaire will collect detailed information about the training received by Water Minders.

[9] See particularly the DRWS Procedure Manual, dated December 2003 and the Village Liaison Officer Village Journal, undated. These set out in some detail the Project Life Cycle and the procedures to be followed in each phase.

[10] See Section 4.7 for a discussion of proximity threats to internal validity from contamination.

The remaining 150 villages did not constitute part of the evaluation sample, and NORC did not track progress or collect data from these villages. Annex B provides a map of the treatment and control villages used in the present study.

Table 1 below presents construction start and end dates, and PHAST community training dates for villages in Phases A^{rev}, A_1, and C. The *official construction completion date* represents the point at which all construction activities of water and sanitation structures (water systems and VIP latrines) were completed, inspected and certified; on this date, after all defects and problems have been rectified by the contractor, the engineer issues the village a Certificate of Completion (CoC). Table 1 also presents the *Approximate physical construction completion date prior to inspection*, which represents the month and year in which all physical construction activities were completed; although some defects had yet to be rectified, it is reasonable to assume that this is the date on which the water systems became operational in each village. For the purpose of the evaluation, it seems appropriate to use the latter, rather than former, date as the point at which treatment commenced in a given village.

The construction of water systems in the 50 Phase A treatment villages commenced between December 2010 and March 2011. Although they were scheduled to be completed by September 2011, there were several delays in the construction schedule. Most importantly, in April 2012, 13 to 16 months after construction began, MCA terminated three contracts of the construction companies responsible for building the water systems in 11 Phase A treatment villages. One year later, in April 2013, a new contractor took over in these 11 treatment sites, and continued the interrupted construction process. These 11 villages constitute the majority of Phase A_1[11]. For ease of presentation and differentiation, we will refer to the group of 39 villages that continued construction with no contractual disruptions as Phase-A^{rev} villages.

Construction of Phase C control villages commenced between January and April 2013.

The construction of water systems in all but five of the 39 Phase A^{rev} treatment villages was physically completed between June 2011 and March 2012, which indicates a minimum 10 months treatment period before construction began in Phase-C villages. Three others were completed in the last quarter of 2012, and two were delayed until September 2013. Completion of construction in all Phase-A_1 villages was severely delayed with nine completing physical construction in Jul-Nov 2013, and two in February 2014. Hence, of the group of 50 treatment villages, construction in all 11 Phase-A_1 villages and one Phase A^{rev} was completed only after construction in Phase-C villages had already begun.

[11] In DRWS and MCA documentation, the original treatment group is referred to as "Phase A." However, following the splintering off of 11 Phase-A_1 villages, the original treatment group *minus* the Phase-A_1 villages was also referred to as "Phase A". In this revised EDR, to avoid confusion, we refer to the original 50 Phase-A villages as Phase A and the reduced set of 39 Phase-A villages (the set minus the delayed Phase-A_1 villages) as Phase A^{rev}.

Table 1: Construction and Training Dates for Phase A^{rev}, A_1, and C Villages

			Dates (number of villages)				
Phase	Type	# villages	Start of construction*	Approximate completion of physical construction, prior to inspection*	Official construction end date (CoC issued)*	PHAST community training†	VWHC training†
A^{rev}	Treatment (T1)	39	Dec 2010 – Feb 2011 (39)	Jun 2011 – Mar 2012 (34) Oct 2012 – Dec 2012 (3) Sep 2013 (2)	Oct 2011 (2) Dec 2011 – Feb 2012 (19) Aug 2012 (13) Sep 2013 (5)	Feb 2008 (1) Apr 2008 (1) Sep– Dec 2008 (26) Mar 2009 – Nov 2009 (10) Nov 2010 (1) May 2013 (1)	Jan 2010 – Nov 2010 (37) May 2013 (2)
A_1	Treatment (T2)	11	Jan 2011 – Mar 2011 (11)	Jul 2013 – Nov 2013 (9) Feb 2014 (2)	Sep 2013 (4) Jan 2014 (4) Mar 2014 (1) Aug 2014 (2)	Apr 2008 – Dec 2008 (7) Apr 2009 (2) Nov 2010 (2)	Jul 2010 (1) Oct 2010 – Nov 2010 (10)
C	Control	48	Jan 2013 – Apr 2013 (48)	Mar 2013 (1) May 2013 – Jun 2013 (7) Aug 2013 – Dec 2013 (39) *Unknown (1)*	Sep 2013 – Mar 2014 (46) *Unknown (2)*	Mar 2009 (1) Jan 2010 – Oct 2010 (17) Mar 2011 (1) Nov 2011 (3) Feb 2012 – Nov 2012 (25) Mar 2013 (1)	Jan 2012 (1) Mar 2012 (7) Oct 2012 – Jul 2013 (39) Nov 2013 (1)

Sources: * MCC spreadsheet, data from Cowater monthly reports
† Cowater / NORC

Early on in the Compact, the Government of Lesotho supplemented MCA investments to enable every household in a treatment village to receive a VIP latrine. The 100-percent coverage plans were based on listings of households within the village conducted by DRWS. For various reasons – quality issues during the listing, which resulted in some households being missed, and a lag between listing and start of construction, during which new households were built – some households in Phase-A villages did not receive VIP latrines. DRWS attempted to rectify this problem in Phase C, by adding new households to the listing immediately prior to construction.

While we do not have any empirical evidence as to the number of households that did not receive VIP latrines in Phase-A villages due to the aforementioned reasons, there is reason to believe that is not likely to have been more than 2 percent of the number of households in a village over the period.[12] Given that the average probability of household selection into our sample was 13 percent, this means that just a little more than one-quarter of one percent of our sample may have included such households. We propose to deal with this anomaly by simply dropping these partially treated Phase-A households from our sample.[13] Moreover, if an endline survey is administered we will ask households in control villages when they moved into the village and drop those households that have resided in the village for less than or equal to the time that led us to drop Phase-A households. This ensures that both treatment and control groups remain in balance statistically.

A second consequence of the above implementation plan is that there is no way for the evaluation to disentangle the separate contributions of the water system and VIP access using a purely design-based approach. Since both treatment and control villages also received PHAST training prior to the baseline survey, this means that the design also cannot infer the contribution of PHAST to any outcomes.

Since implementation of the treatment is phased in over time, we expect the impact will similarly be realized in phases. We anticipate that there will be a lag in time between service becoming available or instruction provision (e.g., hygiene promotion) and the resulting improved hygiene practice or higher-level outcome (e.g., reduced incidence of water-related disease). Impacts are not likely to be instantaneous and some may take considerable time to be fully realized. Therefore, an appropriate evaluation of benefits from improved water and sanitation employs a long time horizon and takes into consideration the time lag between receipt of treatment (the intervention) and the intended result.[14] These anticipated lags between interventions and their effects, shown in Table 2, have guided the timing of treatment phase-in designed into the (pipeline) evaluation as well as have dictated when the various surveys associated with this evaluation take place.

[12] In particular, if one takes rural population growth as an approximate indicator of the increase in households in Phase-A villages then, according to the World Bank, rural population growth in Lesotho grew at 1%, 0.8% and 0.0% for the years 1990, 2000, 2010. Hence, 2% would seem like a conservative value to use as a worst-case scenario. [The World Bank statistic is taken from http://www.tradingeconomics.com/lesotho/rural-population-percent-of-total-population-wb-data.html on 5 January 2015.]

[13] We say "partially treated" because these households benefit from the water system, but do not have a VIP and, so, did not receive the full sanitation intervention.

[14] http://www.who.int/water_sanitation_health/wsh0404summary/en/

Table 2: Minimum expected time required for detection of treatment effects		
Treatment*	**Time lag after treatment**	**Impact detected***
Availability of VIP latrines and safe drinking water closer to the household	3 months	Time saved
Access to safe drinking water, improved sanitation, and better hygiene practices	6 months	A reduction of WRD
Access to safe drinking water, improved sanitation, and better hygiene practices	9 months	Improvements in productivity and income due to time saved

*Note that "better hygiene practices" are hypothesized to be due to PHAST training. This training occurred *before* baseline and, therefore, its influence cannot be detected under a purely design-based evaluation. See the text for details.

With the exception of two, all treatment villages received PHAST early on in the project, prior to the start of construction activities in a village. However, this training was not synchronized with the construction sequence of the water projects in treatment and control villages. Thus, its roll-out was not subject to randomization. For example, as is evident in Table 1, PHAST in several control villages (Phase-C) occurred in 2010, as much as 12-15 months before the water and sanitation treatments were in place in Phase-A^{rev} villages. Therefore, though MCC may consider the hygiene awareness activities as a separate component of the rural water activity under the program, from the evaluation design point of view it is not a different treatment, since all experiment units received it. On the other hand, since the timing of PHAST implementation varied by village (as shown in Table 1), this *may* allow NORC to assess (statistically identify) using a model-based technique the contributory (i.e., interaction) effect of hygiene awareness on the water and sanitation components of the program.

2.2 Program Logic for the Rural Water Project

MCC's new program logic diagram (see Annex A) for the Lesotho rural Water and Sanitation Project, shared with NORC in March 2013, clearly presents activities and outputs that are linked to three levels of outcomes: short-term, intermediate and long-term.[15] These effects are:

Short-term outcomes

- Increased hygiene awareness among communities
- Increased access to improved sanitation
- Increased access to improved water sources
- Increased awareness/knowledge of Water Committees, Water Minders, and communities in maintaining systems

Intermediate outcomes

- Improved hygiene behavior[16]

[15] While NORC reviewed the ERR in addition to the program logic and the literature when establishing the research hypotheses with MCC, NORC was informed by Jennifer Sturdy (DPE/EE-ME) in an email of April 23, 2014 that she had spoken with MCC's economist for Lesotho, Sarah Olmstead, and "she confirms that for this Compact and Rural Water specifically, there is no need to link to the ERR and Beneficiary Analysis."

[16] An MCC reviewer stated that "Improved hygiene behavior is too broad to be an outcome – what is intended here? Hand Washing? Chlorine treatment for water?" This outcome, however, is taken directly from the MCC Program Logic, which includes "Improved hygiene behavior" as an Intermediate Outcome.

- Decreased water-related illness
- Reduced expenditure on medical care
- Time saved in water collection
- Maintenance of systems by Water Minders

Long-term outcomes

- Increased productive activity (productivity)
- Increased income

These outcomes are taken directly from MCC's Program Logic diagram for the Rural Water and Sanitation Project.

2.3 Evaluation Hypotheses and Indicators

The evaluation hypotheses for the impact evaluation are directly linked to the outcomes presented above and in MCC's program logic. Table 3 presents the following information:

- Maps out the evaluation hypotheses related to the rural water-supply investments to key outcome/impact indicators (Columns 1 and 2);
- Indicates the minimum time of exposure necessary to detect changes in the outcome indicators (Column 2, in parentheses);
- Maps outcome indicators to treatment indicators, as shown in the pathways in MCC's Program Logic (Column 4)
- Presents data sources for outcome and treatment indicators (Columns 3 and 5)
- Indicates the analysis method proposed to measure impacts (Column 6)

In addition to the hypotheses in the table, the evaluation plans to examine hypotheses related joint treatment effects (e.g., PHAST plus latrine). Combinations of treatments lead to a further set of hypotheses but are not separately included in the table (other than with the phrase, "singly and jointly") to minimize complexity.[17] As described, below, such tests require appending an estimated model (continuous-treatment-variable (CTV) attribution equation) to the primary design-based (experimental) approach. Since NORC did not have control over the exact timing of intervention activities, this aspect of analysis was not foreseen in the original sampling design and it not certain whether strata sample sizes will be sufficient to detect these contributory effects.

[17] Such additional hypotheses are simply combinations of the primary treatment, namely, training, VIPs, and water systems.

Table 3: Rural Water Project: Mapping of Evaluation hypotheses to Impact Indicators, Treatment Indicators and Data Sources					
	Outcome or Impact Indicator		Treatment Indicator (variable)		
Evaluation period / Hypothesis	Description (Months required)	Data sources	Description	Data sources	Analysis method
Short-term Outcomes					
1. Participation in PHAST training improves hygiene awareness in communities	Degree of household hygiene awareness (3 months)[a]	IEMS HALEH[b]	Measure of household participation in PHAST training(s)	IEMS HALEH	CTV
2. Access to improved water systems increases household use of safe drinking water	Degree to which household uses safe drinking water (3 months)[18]	IEMS	Water system constructed System reliability measure	Cowater Monthly Progress Reports VWHC[c] questionnaire	RCT
3. Installation of a VIP increases use of improved sanitation	Frequency of household use of VIP latrine (3 months)	IEMS	VIP latrine constructed	Cowater Monthly Progress Reports	RCT
4. Access to improved water source reduces time spent collecting water/washing clothes[d]	Time (minutes) spent collecting water[e] Time spent washing clothes (by economically active HH members, women) (3 months)	IEMS	Water system constructed System reliability measure	Cowater Monthly Progress Reports VWHC questionnaire	RCT
5. Greater hygiene awareness leads to improved hygiene behavior	Degree of hygiene behavior (3-6 months)	IEMS HALEH	Degree of household hygiene awareness	IEMS HALEH	CTV
Intermediate Outcomes					
6. Increased use of safe drinking water and improved sanitation/hygiene behavior singly and jointly reduce incidence of WRD[f].	Incidence of diarhhea over last 2 weeks - in children, women, economically active adults (6 months)	IEMS	Degree household uses safe drinking water Degree of household hygiene awareness Frequency of VIP latrine use	IEMS HALEH	RCT, CTV
7. Increased use of safe drinking water and improved sanitation/hygiene behavior singly	Household medical expenditures (6-9 months)	IEMS	Degree household uses safe drinking water	IEMS HALEH	RCT, CTV

[18] Within the current SOW of the IEMS, NORC does not plan to test water quality. However, water quality measures would constitute a far more accurate measure of "safe drinking water." DRWS does collect measures of water quality, which NORC has had access to in the past. If MCC authorizes us to expend additional effort trying to access this secondary data for the relevant time periods and sites, we will do and, depending on its suitability, we will incorporate it into the evaluation.

Table 3: Rural Water Project: Mapping of Evaluation hypotheses to Impact Indicators, Treatment Indicators and Data Sources					
	Outcome or Impact Indicator		Treatment Indicator (variable)		
Evaluation period / Hypothesis	Description (Months required)	Data sources	Description	Data sources	Analysis method
and jointly reduce expenditure on medical care			Rating of household hygiene awareness Frequency of household use of VIP latrine		
8. Increased knowledge about water system maintenance among WMs, VWHC, and households increases water system reliability and water quality [g]	Number of interruptions in supply (system reliability) Water quality (6 months)	DRWS water quality surveys	Training of WM and VWHC completed[h] Rating of WM knowledge of system maintenance Rating of community awareness Maintenance compliance or other source	WM[i] question-naire IEMS VWHC question-naire Maintenance compliance report[j]	CTV
9. Reduction in household's time collecting water and washing clothes increases school attendance	# of days that school-age chil-dren in household attend school in past two months (6 months)	IEMS	Time (minutes) spent collecting water Time (minutes) spent wa-shing clothes	IEMS	RCT
10. Reduction in WRD will pro-vide time-savings to house-holds	Time (days) lost from illness Time (days) lost in caring for the sick (6 months)	IEMS	Incidence of WRD	IEMS	CTV
Long-term Outcomes[l]					
11. Reduction in time spent col-lecting water/washing clothes increases women's produc-tive activity and income	Work hours, Income (of women in household) (9 months)	IEMS	Time (minutes) spent collecting water Time (minutes) spent wa-shing clothes	IEMS	CTV
12. Time saved from reduced incidence of WRD will lead to increased household income	Hours worked Income (of household as a whole & in-dividual members) (9 months)	IEMS	Time (days) lost from illness Time (days) lost in caring for the sick (for economically active household members, by gender)	IEMS	CTV

Table 3: Rural Water Project: Mapping of Evaluation hypotheses to Impact Indicators, Treatment Indicators and Data Sources					
	Outcome or Impact Indicator		Treatment Indicator (variable)		
Evaluation period / Hypothesis	Description (Months required)	Data sources	Description	Data sources	Analysis method
13. Access to improved water system, VIP, PHAST singly and jointly lead to increased household income	Hours worked Income (of household as a whole & individual members) (9 months)	IEMS	Water system constructed Household participation in PHAST training(s) VIP installed in household	VWHS questionnaire HALEH IEMS Cowater Monthly Reports	RCT, CTV

(a) The AMP questionnaires, as initially created by DRWS, consisted of a simple checklists of 10 items to be rated as adequate/inadequate. At that juncture, NORC and Cowater envisioned developing a set of simple rating corresponding to the level of effort in terms of committee performance, Water Minder capabilities, use of latrines, etc. The current AMP instruments, developed by NORC and Cowater, have evolved well beyond such a checklist. However, we still plan to compress these questions into 5-10 areas most relevant to the outcomes of interest and rank responses on a scale vis-à-vis the outcome being measured (degree or rating of hygiene awareness, hygiene behavior). We will define the ranking and scoring guidelines as we finalize the AMP instruments and prepare for the endline data collection.
(b) Health Latrine and Environmental Health questionnaire
(c) Village Water Health Committee questionnaire
(d) Although "time saved in water collection" is presented as an intermediate outcome in MCC's Program Logic, we anticipate changes to occur in a 3 month time frame. As such, we include it here as a short-term outcome.
(e) *Comment from MCC reviewer*: Please make sure we collect data on the location of alternative sources and the measured as opposed to reported time savings. *NORC Response*: The IEMS questionnaire includes a set of detailed questions about primary and secondary sources of water, including time and distance to alternative sources, as reported by respondents. This data was collected at baseline and midline. However, we did not independently measure the time to each water source; therefore it will not be possible to capture "measured, as opposed to reported, time savings." If this is of interest to MCC, we could establish the distances to the primary water sources through geographic information systems (GIS) and from a coefficient of distance to time, a fairly accurate measure of the time and time saved as well. However, this was not planned for in the evaluation design.
(f) *Comment from MCC Reviewer:* In particular (a) are water supply and health benefits related or not - could the programs be separated with benefit loss? and (b) if they are linked then how are they linked - do health benefits also diminish as a function of distance from a source similar to time saving benefits? *NORC Response:* The MCC program logic and hence, our evaluation is primarily concerned with measuring health benefits associated with improved water supply, hygiene, and sanitation. Research does not show that health benefits diminish as a function of distance from the improved water source. Health benefits are usually associated with the separate and combined effects of safe water supply, improved sanitation, and hygiene practices (hand-washing at key times) rather than distance from water source. In fact, distance to water source has not been found to have differential impact on health except where there is zero distance i.e. piped water to the house or internal plumbing.
If separated, the water and health programs would be lead to some benefit loss. The PHAST messages e.g. hand-washing and use of Oral Rehydration Therapy (ORT) to treat diarrhea, could be considered as both health and water interventions. The link between water supply and health benefits are, as stated above, the result of widely acknowledged improved hygiene practices.
(g) The Program Logic also indicates that this process is strengthened by the Aftercare Trainings.
(h) Includes Aftercare Training.
(i) Water Minder questionnaire.
(j) If available from DRWS

(k) *Comment from MCC reviewer*: Employment or entrepreneurial options are very limited in rural areas, especially in remote villages with limited access to transportation, let alone markets. That being said, the probability of these outcomes materializing is perhaps low unless another stimulus is introduced into this scenario and capitalizes on intermediate outcomes, which are more realistic. *NORC Response:* These long term outcomes – productive activity and income – are taken directly from MCC's Program Logic and, as such, we have opted to include them in the revised Table 3.

3. LITERATURE REVIEW

3.1 Summary of Existing Evidence

The benefits of WASH (Water, Sanitation, and Hygiene) programs are often cited. Meta-analysis and systematic reviews (such as Fewtrell *et al.*, 2005) found water, sanitation, and hygiene interventions to reduce significantly the risks of illness such as diarrhea illness. In terms of the benefits of improved water quality specifically, there is wide consensus in the research of the positive and significant health benefits, in both meta-analysis and systematic review (Esrey *et. al.*, 1999) and in relevant studies in rural areas (see Annex C). Safe drinking water improves health largely by reducing occurrence of diarrhea, a very common illness in the developing world, and other water-related illness. The largest health gains, especially in terms of mortality, are to children under five (see Annex C).

In regard to the hygiene and sanitation component of the program, there is also evidence in the literature of health benefits of these such programs, and there is some evidence that all of the WASH interventions more effective when combined. Improved health, in turn, should lead to a number of benefits, including reduced medical costs, reduced time seeking medical care (which can therefore can lead to more time spent at productive income-generating activity), and improved productivity (which should lead to improved wages or outputs per hour).

The literature also indicates that improved access to water reduces water collection time, releasing time and resources for productive activities, such as work and school. However, the data on the amount of time saved is scarcer. Nonetheless, we highlight some in Annex C. We also highlight the literature on the longer term impacts of the program, such as increased productivity, school attendance, and ultimately, income. This is an area, in particular, where the MCC Lesotho Rural WSP Evaluation could make a contribution to the literature.

3.2 Gaps in Literature

We found numerous studies that demonstrated that improved water enhanced health in proportion to its use (i.e., water programs were most effective if other unimproved sources were not continued to be used). However, we were not able to locate literature on how the *duration* of improved water access contributes to health outcomes (Hypothesis 7). This is an area where the MCC Lesotho Rural WSP Evaluation can make a unique contribution to the literature. In addition, although we found many studies on the benefits of hygiene education, there does not appear to be a literature on the how the intensity (i.e. quality, duration) of such training affects outcomes, and this is another area where our study can make a contribution to the body of research.

Finally, most studies conceptualize water supply as either "improved" or "unimproved." However, despite a water intervention, unimproved water sources may still be used to some degree, especially if the "improved" water source is not maintained over time and proves to be unreliable.[19] Likewise, delays in program implementation may have resulted in delayed access and

[19] Hence, the literatures and the Program Logic, stress the importance of maintenance/sustainability of water supply improvements.

behavioral change related to improved water sources than was originally anticipated. Our evaluation addresses this through its unique design: in addition to the simple treatment/control classification used in the difference-in-difference approach, it also proposes the use of a CTV or dose-response design, whereby we measure the *degree* to which households use improved water, in order to more accurately measure different levels of water access to program households.[20,21]

The impact evaluation for the rural WSP in Lesotho will seek to determine to what degree the intervention's improved water, sanitation, and hygiene reduce water collection time and water-related illnesses, thereby creating benefits to the target population such as improved time use, productivity, and incomes. In doing so, the study will also contribute uniquely to the WASH impact literature by utilizing wide-ranging data sources that examine outcomes based on more comprehensive indicators (including unique rarely used indicators, highlighted above), and employing a dual-design methodology that allows a more nuanced view of this type of WASH intervention.

4. EVALUATION DESIGN

The evaluation has policy relevance to national and international instruments and strategies and provides a focus on income generation, the long-term maintenance of water systems, quality of life improvements, and progress towards goals found in Lesotho's National Development Plan.

The promotion of increased productivity, economic growth and the reduction of poverty is the overall rationale of MCC's intervention in Lesotho.[22] While the effect of the present interventions on waterborne diseases is well documented, the intricate causal links between the rural intervention and rural incomes are not well established in the developmental literature. This evaluation will help establish these pathways.

International concern in the water and sanitation sector now focuses as much on the sustainability as on the delivery of water systems.[23] In drawing out the lessons for the water sector, the End of Program Review also emphasizes the need to address sustainability of rural water systems in its financial, institutional capacity and organizational dimensions.[24,25] This evaluation will, however, only address sustainability at the level of functionality, by assessing the effects of

[20] By different levels of water access we mean distances to alternative water sources available to the household.

[21] *Comment from MCC Reviewer:* I'm not clear on the definition of "different levels of water access" when the system was designed to make sure that the maximum distance between any given household and the water system was 150m. Also, access was equal to all folks once the system came online. *NORC Response:* This comment contradicts a previous comment by another MCC reviewer that stated, "I'd like an estimate of the probability of using an improved source as a function of distance to new/old source, cost of new/old source, other variables. This will allow us to explore the intrinsic value of value placed on an improved water source." In our analysis, we will examine the effect of distance to the improved water sources.

[22] MCC Lesotho Compact

[23] Water Council. What are the benefits of safe water supply and sanitation? http://www.worldwatercouncil.org/library/archives/water-supply-sanitation/

[24] MCA, Lesotho. January 2014. End of Program Review: Water and Sanitation Sector, Volume 2. 5.2 Lessons for the water sector.

[25] A MCC reviewer concurred with this statement, stating the following: "This, by the way, is a major concern major concern on the DRWS projects. On the sanitation side, it is very likely that the VIP may only have an average service life of five to seven years

VWHS and Water Minder training on the ongoing operation and maintenance of water and sanitation systems (see Post-Construction Phase of MCC's Program Logic).[26]

The basic motivation for water and sanitation interventions is increasing the efficiency of water-intensive household activities, improving health, increasing productive time and productivity of household members and, ultimately, increasing incomes and improving quality of life. Finally there is the overarching commitment to meeting the Millennium Development Goals in Lesotho's the water sector. The latest report of the UNDP on Lesotho mentions progress being made generally but spells out the technical difficulties and investment challenges associated with rural water supply and the need to "begin to shift from urban to rural provision". Similarly it mentions the "pronounced" challenge in rural sanitation if an all-rounded advance is to be made.[27] Both these policy priorities will be fully examined in the evaluation of the Lesotho Rural WSP.

4.2 Overview of Quantitative and Qualitative Evaluation Design

The impact evaluation design of the rural water intervention employs two separate approaches so as to ensure a "defensive evaluation"—that is, so that results can be obtained in spite of unforeseen events or mishaps. One is a phased randomized control trial (RCT) in which 100 water systems were allocated (roughly) equally to treatment (i.e., early treatment) and control (i.e., late treatment) groups within each of the ten districts of Lesotho.[28] The other is a continuous treatment-variable (CTV) model in which the intensity of treatment can be estimated without the need to isolate a separate control group. As we shall see, while our revised design provides techniques to overcome the threats to internal validity that delayed implementation created for the RCT design, there are compelling reasons to give a greater role to the CTV approach as well as to carry out endline surveys.

The original evaluation design, developed under NORC's first contract with MCC, focused on a randomized design, under which NORC, with MCA planned for a 6-9 month gap between the end of construction and rehabilitation of treatment water projects (in 50 Phase-A villages) and the start of water projects in the control areas (in 50 Phase-C villages) to ensure sufficient time for impact to be realized in all outcomes of interest. A nine month gap would allow us to measure long-term outcomes such as changes in productivity and income, while a six-month gap would limit the impact analysis to short-term and intermediate outcomes. Under the original design, all Phase-A villages had a largely similar construction timeline with concurrent start and end dates of construction; thus, it was reasonable to expect that there would be a nine-month (or, at a

[26] A MCC reviewer pointed out in the EDR review that "More critical is the lack of budgetary support for DRWS' O&M operations. For instance, lack of spare and replacement parts at the district level results in water system being inoperable and in some cases, this creates a problem far worse than before these systems were installed." While we agree with this view, as it is currently envisioned, our evaluation does not look at the institutional, financial, and structural aspects of sustainability, except for instances when this information may be elicited through the HALEH, VWHC, and WM questionnaires to be administered at endline.

[27] See detailed comments on water and sanitation targets in the 2011 Report on Lesotho. http://www.undp.org.ls/millennium/Millenium Development Goal 7.pdf

[28] See Section 2.1.3 on project implementation for the details on how treated villages were randomly selected.

minimum, a six-month) lag between the end of construction of the 50 Phase-A villages and the start of construction of the 50 Phase-C villages.[29]

Delays in the construction of Phase-A (early treatment) water systems, resulted in 11 treatment villages (Phase-A_1 villages) undergoing construction concurrently with the Phase-C control villages. This overlap has called into question the validity of the original evaluation design—and the efficacy of administering an endline.[30] As Table 1 demonstrates, construction was completed in only 70 percent of Phase-A villages (34 of the 50) nine months before construction commenced in Phase-C control villages in January 2013. For these 34 villages – which belong to the Phase-A^{rev} group – the time of exposure to treatment before controls began receiving treatment between in January 2013 range from 10 to 19 months. Since the midline data collection preceded the start of construction in Phase-C villages (i.e., November to December, 2012), duration of exposure to treatment by the time of midline data collection for these 34 Phase-A^{rev} villages was about 8-17 months.[31] Five villages in Phase A^{rev} and all 11 Phase-A_1 villages had not been exposed to treatment (i.e. construction of their water systems had not been completed) by the time of midline data collection. The reduction of the treatment sample at midline by 30 percent has implications for the statistical power of the evaluation design to detect impacts at the level generally desired by MCC.[32]

In what follows, we discuss the implications of these construction delays for the evaluation design and indicate what evaluation questions can and cannot be addressed without an endline. These consequences are brought together in Section 4.3.4, which discusses the pros and cons of alternative design strategies to deal with the unforeseen complications in implementation.

4.3 Impact Evaluation Design and Methodology

This subsection describes the two separate methods we propose to use for the impact evaluation of the rural water intervention and where they would be applied.

4.3.1 Design using double-differenced cluster-randomized control trials

For hypotheses in Table 3 to be addressed by an experimental (i.e., design-based) approach, we will use a difference-in-differences estimator to infer the effect of the intervention on outcomes of interest. These outcomes relate primarily to households (e.g., time saving), but also to individuals (e.g., incidence of diarrhea). The original design was predicated on the fact that Phase-A treatment villages (what are now Phase-A^{rev} and Phase-A_1 villages taken as a group) and Phase-C control villages were statistically identical by design so any systematic differences between

[29] The requisite time lag to detect effects is measured from the end of construction in treatment group to start of construction in control group (rather than end of construction in control group) because the control conditions change even with the inception of construction/rehabilitation. For example, during the construction period, water supply is interrupted, VIPs in some household are completed and become operational, and the perceptions and attitudes of household members are affected. The end-of-construction in treatment to start-of-construction in control group time frame allows us to avoid such contamination and preserve a largely untouched control group.

[30] We examine the implications for an endline, below, in Section 4.3.4.

[31] Five villages in Phase A^{rev} had no exposure to treatment at midline data collection, because completion of construction coincided with or surpassed the midline data collection.

[32] For clarity and to avoid confusion with other similar percentages in this report, note that the 30-percent reduction is calculated from [(4 Phase-A^{rev} villages) + (11 Phase-A_1 villages)]/(48 total villages to treat).

them at the end of the six- or the nine-month treatment period would be due solely to the MCC intervention.[33] This approach allows the research team to causally attribute the detected change in the outcome indicators to the MCC intervention activities.

For didactic reasons, we proceed in two steps. First we lay out the difference-in-differences model developed for the original evaluation design (i.e., where there was only one treatment group). Then, we show how that model needs to be modified to account for the delay in treatment to a Phase-A_1 subgroup of the original Phase-A villages.

The Basic Difference-in-Differences Model

The difference-in-differences attribution model for each hypothesis evaluated includes an indicator variable for whether a household or individual was treated, a variable for time of the survey (i.e., baseline, midline, or endline), an interaction term of these two variables, and covariates and fixed effects to capture other influences on the outcome. The coefficient of the interaction term is the difference-in-differences estimator that measures the effect of the intervention.

Mathematically, the attribution equation to estimate would look like:

$$Y_{it} = \beta_0 I_i + \beta_1 R_t + \beta_2 I_i R_t + \boldsymbol{X}'_{\boldsymbol{it}}\boldsymbol{\delta} + \{\boldsymbol{D}_{\boldsymbol{it}}\} + \{\boldsymbol{H}_{\boldsymbol{i}}\} + \varepsilon_{it} \tag{1}$$

where Y_{it} is an outcome measure of household i in round t; I_i indicates the household's village's intervention status ($I_i = 0$, if household i is from a control village and $I_i = 1$, if household i is from a treatment village);[34] $\boldsymbol{X}'_{\boldsymbol{it}}$ is a vector of time-varying characteristics (such as household size age, and education of the head of household); R_t is the round dummy ($R_t = 0$ at interval start, t=0 and $R_t = 1$ at interval finish, t=1), $\{\boldsymbol{D}_{\boldsymbol{it}}\}$ is a vector of interaction terms between village dummies and round, $\{\boldsymbol{H}_{\boldsymbol{i}}\}$ is a vector of (absorbed) household fixed effects, ε_{it} is an error term and the β_j and $\boldsymbol{\delta}$ are parameters to be estimated.[35] The estimated value of β_2 captures the effect of treatment.

Correcting for Systematic Bias

As already mentioned, Phase-C villages may not be the correct counterfactual for Phase-A^{rev} villages because the former was constructed to include control villages for Phase A, that is, Phase-A^{rev} *and* Phase-A_1 villages. Recall that Phase-A_1 villages had their water system construction systematically delayed. If the reason for the delay is in any way correlated to how households are impacted by the MCC intervention, then the difference-in-difference analysis laid out, above,

[33] One reviewer from MCC suggested that NORC replace the eleven Phase-A_1 villages in the treatment group with (a presumably random set of) other treated Phase-A villages. The problem with this suggestion is that no baseline (or midline, for that matter) data were collected from such other villages.

[34] To respond to a question by one MCC reviewer, according to the implementation design, all households in a village targeted for treatment were treated, including receipt of a VIP latrine. The degree of treatment in the case of use of the new water system or PHAST training is better analyzed in the CTV model, below. The RCT analysis here – and the variable, I – therefore, are intended to detect the effect of "treatment of the treated". Endogeneity is handled using instrumental variables and is addressed, below.

[35] Note that (1) the objective behind including village and round interaction terms is to try to capture any major contextual change at the village level that could affect the outcomes of interest and (2) by including household fixed effects any characteristics that may confound the treatment effect are isolated (controlled), as long as the characteristics are time-invariant..

would not have internal validity. Hence, before conducting above analysis, we must determine whether we can detect any systematic observable statistical differences (lack of balance or common support) between Phase A^{rev} (treatment) households and Phase C (control) households at baseline.[36] If no statistical differences are detected then we may simply proceed with the model specified, above; if there are differences, then there are three available options we plan to carry out and then compare for robustness.[37]

To present these options the mechanisms of the pipeline approach must be laid out a bit more technically. This approach uses beneficiaries already chosen to participate in a project at a later stage (Phase C villages) as the control group for those chosen to participate at the start (Phase A villages). Through successive survey rounds data are collected on treatment and control groups under different durations or even features of the intervention. The assumption is that treatment exposure is both cross-sectionally and temporally assigned randomly so villages selected to receive the intervention in the future are similar to those selected to receive it earlier; villages for these groups, therefore, should be comparable for outcome variables of interest. Such comparisons are feasible since the timing of these phased interventions is known at a village level from administrative data and also (but less precisely) at a household level from the repeated surveys.

The basic mechanics of the original two-period pipeline approach is illustrated on the right side of Figure 1, in which three rounds of a survey are conducted, a baseline, a midline, and an endline.[38] (As we shall see, the pipeline design may also be viewed as illustrating two distinct treatments or one treatment applied twice.) The left-hand-side of Figure 1 illustrates the *revised* design required due to the existence of Phase-A_1 villages. Critically, we see that due to the implementation delay, the original control group (C_C in the figure) must now be thought of, at least in principle, as comprising two parts, Phase-C_2 villages that act as the counterfactual for the on-time Phase-A^{rev} villages and Phase-C_1 villages that act as the counterfactual for the delayed Phase-A_1 villages.

Using the nomenclature of Figure 1, the first and simplest correction for selection bias caused by the implementation delay is to calculate the changes in the appropriate control subgroup over the evaluation period using the information we have for C_C, C_1, and C_2. This is done by subtracting the (appropriately weighted) contribution of the average differences of C_1 villages from those of C_C villages. This should result in the appropriate residual counterfactual control, C_2, for the Phase-A^{rev} villages (i.e., Phase-A villages without Phase-A_1 villages) in Period 1. In other words, if N_C, N_1, and N_2 are the number of households in Phase C, Phase A_1, and Phase A^{rev} then for any variable, $z_{i,t}$, where $\Delta z_{i,t} = z_{i,t} - z_{i,t-1}$,

[36] A preferred and more stringent test, which we will also conduct, would be whether Phase-A and Phase-A_1 villages share a common support. We do not mention this in the text because it is possible that there are too few Phase-A_1 villages to reject a common support at a satisfactory level of significance.

[37] At MCC's request, NORC held off on conducting an analysis of the baseline data collected under the first contract with MCC. Because of the evolving thinking around the Lesotho health sector evaluation at MCC and the late development of a Program Logic for the rural and urban water projects, NORC was instructed to delay analysis of the baseline data. Most recently, we were instructed to cease all activities on the health sector evaluation and focus all attention on updating the rural water EDR. As such, to date, we have not analyzed the IEMS baseline data. We agree with MCC that this analysis should be a high priority in moving forward.

[38] As of this writing the MCC has not taken a decision on whether to conduct the endline; the arguments for and against one are presented in Section 4.3.4.

$$\sum_{i \in C_C} \Delta z_{i,t} = \sum_{i \in C_1} \Delta z_{i,t} + \sum_{i \in C_2} \Delta z_{i,t} \quad . \tag{2}$$

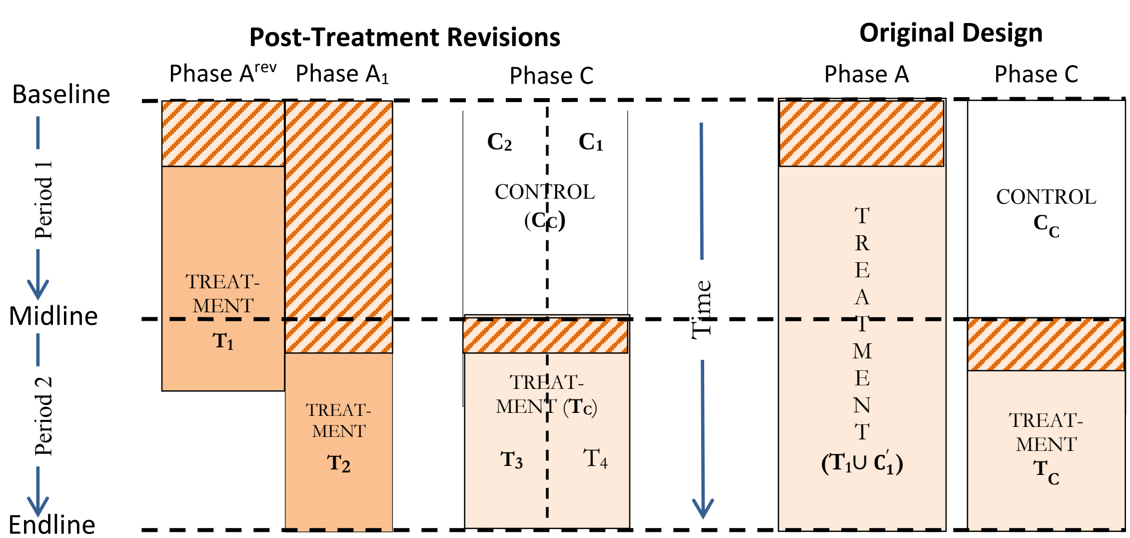
Letting

$$\overline{\Delta z}_{j,t} \equiv \frac{\sum_{i \in \mathrm{C}_j} \Delta z_{i,t}}{N_j} \quad \text{and} \quad \theta_j \equiv \frac{N_j}{N_\mathrm{C}} \quad , \ j = \{\mathrm{C}, 1{,}2\} \tag{3}$$

then it is easy to show that $\overline{\Delta z}_{2,t} = \overline{\Delta z}_{\mathrm{C},t}/\theta_2 - \overline{\Delta z}_{1,t}(\theta_1/\theta_2)$. Thus, the simplest (i.e., non-parametric) difference-in-differences estimate of the treatment effect *over Period 1* is $\overline{\Delta z}_{\mathrm{T}_1,t} - \overline{\Delta z}_{2,t}$, where the first term is the average change in Phase-A^{rev} villages and the second term is the average change in Phase-C_2 villages, the corrected control group for A^{rev} villages.

This simple approach does, however, rest on one testable assumption, namely, that household VIP and water-system-related behavioral changes and their benefits do not begin until the end of construction. This assumption allows us to estimate $\overline{\Delta z}_{1,t}$ using Phase-A_1 households. Without this assumption there would not be an easy way to compute $\overline{\Delta z}_{1,t}$.

Figure 1: Two-step pipeline evaluation design with a retrospective split in treatment



Note: Phase A is the original group of treatment villages, which during implementation got non-randomly split into two subgroups, A^{rev} and A_1. (Phase-C villages were not affected.) Cross-hatched areas indicate that construction was underway but not completed so the benefits of treatment had not yet begun.

The alternative albeit more complicated way we will consider to address the potential selection bias is to create a set of matched comparison villages for Phase-A^{rev} villages from Phase-C villages in the control group.[39] In particular, we model the probability of being "selected" for delayed participation as a propensity score within a probit model. Mathematically, we estimate the probability of village *v* participation, P_v, as $\Pr(P = 1) = \Phi[\boldsymbol{\varphi}'\boldsymbol{Q}_v]$, where Φ stands for the probit distribution, $\boldsymbol{Q}_v$ is a vector of *village* characteristics at baseline and $\boldsymbol{\varphi}$ are parameters to be

[39] Propensity-score models assume selection on observables, i.e., that the selection process was made (and therefore can be revealed to the evaluator) based on criteria (explanatory variables) that the evaluator is able measure and uses in the selection equation that is estimated.

estimated. Estimation of the associated "selection" equation could use such covariates as district ($Q_{1,v}$), village population size ($Q_{2,v}$), distance to urban area ($Q_{3,v}$), average rainfall ($Q_{4,v}$), among others. As an example, $\boldsymbol{\varphi}'\boldsymbol{Q}_v$ might look like:

$$\boldsymbol{\varphi}'\boldsymbol{Q}_v = \varphi_0 + \varphi_1 Q_{1,v} + \varphi_2 Q_{2,v} + \varphi_3 Q_{3,v} + \varphi_4 Q_{4,v} + \xi_v \tag{4}$$

where ξ_v is the error term. Using the estimated parameters of this equation a predicted propensity score value, $\hat{P}_v$, for each village in Phase A^{rev} and in Phase C can be calculated. Using any one of a number of standard propensity-score matching algorithms, one (or more) Phase-C villages can be matched to each Phase-A^{rev} village. The households within this matched set of treated and comparison villages can be used to estimate Equation (1), above, and, thereby, overcome the threat of selection bias (given the assumption in Footnote 39).

A third (and, possibly, best) way to overcome selection bias is to take advantage of the original randomized assignment to treatment and control and construct a proxy variable to block the backdoor channel leading to the bias. We do this by estimating the probability, S_v, of a village v's receipt of treatment (i.e., finding itself in Phase A^{rev}) given it was originally assigned to treatment (i.e., finding itself randomly in Phase A) as $\Pr(S_v = 1) = \Phi[\boldsymbol{\omega}'\boldsymbol{W}_v]$, where now $\Phi[.]$ estimated over the *full* sample of Phase-A and Phase-C villages:

$$\boldsymbol{\omega}'\boldsymbol{W}_v = \omega_0 + \omega_1 W_{1,v} + \omega_2 W_{2,v} + \omega_3 W_{3,v} + \omega_4 W_{4,v} + \zeta_v \tag{5}$$

where W_1 is the original randomization dummy (and so equal to 1 if the village is *either* in Phase A^{rev} *or* in Phase A_1 and equal to 0 if the village is Phase C) and W_k, $k > 1$, are covariates characterizing villages, including reasons for drop-out (i.e., selection into A_1). We then compute $\hat{S}_v$ as a measure of (actual) treatment and add it into the attribution Equation (1), above:

$$Y_{it} = \beta_0 \hat{S}_v + \beta_1 R_t + \beta_2 \hat{S}_v R_t + \boldsymbol{X}'_{it}\boldsymbol{\delta} + \{\boldsymbol{D}_{it}\} + \{\boldsymbol{H}_i\} + \varepsilon_{it} \tag{1'}$$

where $\boldsymbol{X}$ may now include some selection-equation covariates in addition to those proposed for Equation (1) and the appropriate value of S_v for each household i is the one that corresponds to i's village v.[40] The hypothesis test for attribution still focuses on whether β_2 is statistically significantly different from zero.

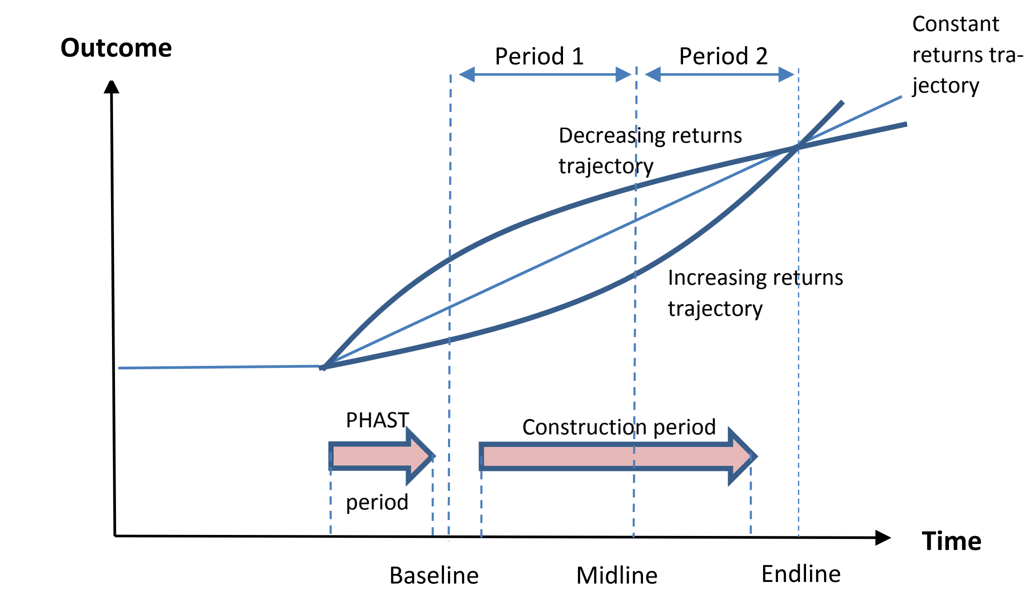
Additional analysis opportunities from a pipeline design

The time differences in exposure to treatment among the various groups should detect, if they exist, statistically significant impacts and help illustrate how the intervention effect accumulates over time, e.g., months using VIP latrines and access to safe drinking water. This should be of significant policy importance since it indicates whether most of the benefits come in the shorter term or whether they only arrive in the longer term. Inferring the evolution (trajectory curvature) of WASH effects is carried out by comparing the size of impact of exposure in Period-2 of those exposed in Period 1 to those only exposed in Period 2. With regard to Figure 1, we need to compare changes over Period 2 (i.e., midline to endline) in the outcome variables for households in T_1 to changes in those variables for households in T_2 over the same period. This is equivalent to

[40] More precisely, each variable in the interaction term of the equation to estimate should first be mean-differenced.

determining whether an outcome is increasing at a constant, increasing, or decreasing rate, as illustrated in Figure 2.

Figure 2: Possible Trajectories of WASH Outcomes



Note: It is unlikely that these trajectories would intersect at the same point, still less at the time of end-line. This figure was constructed in this way precisely to avoid unintended inferences regarding which trajectory might be higher at endline.

To come full circle, using the nomenclature of Figure 1, we see that the formulation in Equation (1) or Equation (1'), depending on the situation, can be used in more than one experiment:

Experiment (Comparison)	R_t	Treatment indicator	Baseline phase of village
Group T_1 to Group C_2	$R_t = 0$ at baseline $R_t = 1$ at midline	$\hat{S}_v$	A^{rev} and C_C
Group T_1 to Group T_3	$R_t = 0$ at midline $R_t = 1$ at endline		

While the detection of accumulating effects is the specific focus of the continuous treatment variable (CTV) approach that follows, a "pipeline" design allows the estimation of such effects through design-based experiments – what MCC refers to as "impact evaluation" – which reduces the spurious biases from unobserved sources of influence. Of course, in this case a comparison of different exposure intervals to treatment requires a longer evaluation period – ideally twice the length of Period 1 in Figure 1, i.e., twelve to eighteen months, depending on the hypothesis. Likewise, another key requirement for the pipeline approach (if you want to assess the returns over time) is the availability of a baseline, midline, *and* endline of data. On the other hand, it is an empirical question whether or not Period 1 is long enough for the CTV approach to detect curvature. We will only know once we carry out the analysis.

4.3.2 Design using continuous-treatment variables

Despite proposing three independent techniques to overcome the design weaknesses of the original RCT, there are two reasons that the evaluation design should not limit itself to a RCT approach.

First, even ignoring the potential for selection bias, the timing of treatments during Period 1 raises a concern for a difference-in-differences approach. That concern is that all members of the treatment group did not receive the treatment at the same time, but rather treatment was staggered over a nine-month period (regardless of whether we define treatment as the beginning of operations or the end of construction). Meanwhile, the midline was only about 9 months after the last villages were treated (barring a few outliers that were treated later). Thus, the difference-in-differences estimator must aggregate households that were treated 9 months prior to the survey with those that were treated perhaps 18 months prior. The resulting average treatment effects, therefore, will be rather blunt and have a distribution with a larger standard deviation than one might want (though the latter might be narrowed with stratification and interaction terms).[41] This is especially problematic if we expect impacts to evolve over time.

Second, there are several hypotheses that, due to the way the intervention was designed, do not submit well to a difference-in-differences approach. For example, this is the case for the tests of the following hypotheses:

> *Hypothesis 1*: Does the greater household participation in PHAST lead to greater household hygiene awareness?
>
> *Hypothesis 5*: Does greater household hygiene awareness lead to improved hygiene behavior?
>
> *Hypothesis 8:* Does greater WM knowledge lead to more reliable water supply?

While in each case there is an experiment one could run in theory, the present evaluation was not set up this way.[42] For example, for Hypothesis 1, above, we did not randomly assign households into groups receiving few PHAST trainings and many PHAST trainings; in Hypothesis 2 we did not randomly assign households into groups of limited household hygiene awareness and significant hygiene awareness; and so on. In these cases no clear control or comparison group can be established (or measured), preventing the application of evaluation designs based on an experiment or quasi-experiment. For these situations, an alternative analytic approach is to estimate a continuous-treatment-variable (CTV) model.[43] Another way to view the CTV model is that it conveniently allows for varying degrees of treatment to be evaluated. For example the intensity of PHAST training (e.g., number of visits, duration of training exposure, or even time since last

[41] Here there is a tradeoff since stratification and interaction terms reduce the sample size for a given hypothesis test.

[42] The original scope of work targeted construction activities. It was only later that NORC was asked to include PHAST, WM, and VHWC and that MCC established the detailed program logic.

[43] An MCC reviewer suggested another application of CTV: "I'd like an estimate of the probability of using an improved [water] source as a function of distance to new/old source, cost of new/old source, other variables. This will allow us to explore the intrinsic value placed on an improved water source." In this application, the dosage is the distance to the water source. The IEMS survey contains a set of detailed questions on the distance (as well as time) to the primary and secondary water sources that would provide variables for this CTV application.

training) can be used as a "dosage". Moreover, this indicator can be "interacted" with other explanatory covariates to infer synergistic effects.

The Basic CTV Model Specification

CTV analysis is based on the specification and estimation of a multivariate regression for each outcome indicator of interest. The dependent (left-hand-side) variable is the change in the outcome—or "response"—indicator of interest (the "effect") and the key independent (right-hand-side) variable is an indicator of the *intensity* of treatment—the "dose". Among the options for measuring treatment intensity include the duration of treatment, frequency of treatment, amount of treatment, time since treatment, and the quality of treatment.[44] Beyond the standard assumptions that must hold for the respective estimation procedure applied (e.g., OLS, GLS) what is required is that the treatment indicator be ordinal or cardinal, multivalued (and, ideally, continuous), and correlated in theory to the response indicator. Each regression will also include covariates to control for other observable factors associated with the changes in the outcome. Unlike the case for the difference-in-differences attribution equation, the sample for the CTV regression is only households in those villages that actually receive treatment. An example of such a model is:

$$\Delta Y_i = a + bT_i^{(1)} + cT_i^{(2)} + dT_i^{(1)}T_i^{(2)} + [g\hat{V}_v + h\hat{H}_i] + \boldsymbol{e}_{\boldsymbol{i}}'(\Delta \boldsymbol{Z}_i) + \varepsilon_i \tag{6}$$

where the T^k are, say, the number of PHAST trainings (including "Aftercare") and the months since the last training, $\Delta\boldsymbol{Z}$ is a vector of household and village covariates, $\hat{V}$ and $\hat{H}$ are village and household propensity-score instruments that we describe, below, and the lower-case letters are the parameters to estimate. Attribution is accepted if the hypothesis that the regression coefficient of the continuous-treatment-variable term(s) is (are) zero can be rejected at a pre-established level of statistical significance, usually five percent; in Equation (6), these coefficients are b, c, and d.

Addressing Threats to the Validity of the CTV Model

The success of the CTV approach, however, depends critically on being able to estimate a properly specified multivariate regression model, i.e., one that addresses selection bias and avoids omitted variable bias (controlling for unobservables). Let us consider each of these threats in turn.

[44] Our household survey will provide information to construct different metrics of diarrhea at each sampling point. We plan to calculate two standard illness metrics (widely used in the field) that are common in studies of diarrheal infections among children in developing countries: (1) incidence (the number of new episodes for each child), and (2) longitudinal prevalence (the number of days of illness divided by total days of observation for each child). We plan to use multiple metrics because they capture different processes. Intuitively, incidence is a measure the rate of transmission of the disease; prevalence estimates the burden of the disease in the population. Epidemiologic theory suggests that incidence is useful for looking at risk factors of disease while prevalence is more useful for measuring disease burden in a population. Longitudinal prevalence of diarrhea is preferred on theoretic grounds and empirically it is more strongly associated with child mortality and weight gain than incidence. (Kleinbaum, D.G., Kupper, L.L., Morgenstern, H. *Epidemiologic Research: Principles and Quantitative Methods.* New York: John Wiley & Sons, Inc., 1982; Morris, S.S., Cousens, S.N., Kirkwood, B.R., Arthur, P., Ross, D.A. "Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence?" *American Journal of Epidemiology* 1996 Sep 15;144(6):582-8.).

As addressed above, there are two primary sources of endogeneity or selection bias, one at the village level and the other at the household level. Limiting ourselves to a Period-1 evaluation (i.e., one with only a baseline and midline), leaves us with a situation in which treatment was withheld from Phase-A_1 villages so it is conceivable that the remaining Phase-A^{rev} village have some (potentially unobservable) characteristics that influence the degree to which our hypotheses of interest are true. Above, we propose two alternative instruments to close this channel of influence, $\hat{S}_v$ and $\hat{P}_v$. Either of these can be considered for V_v in Equation (6).[45]

The second potential source of selection bias is at the household level.[46] For example, some households may have a greater propensity due to unobservable characteristics to participate in a PHAST or Aftercare Training or to take up use of a new water source than other households. If those unobservable characteristics also affect the impacts of treatment then the size or even statistical significance of the parameters in Equation (6) could lose their internal validity. This can be addressed by finding a variable that is correlated to the probability of uptake but uncorrelated to responsiveness to treatment. One such variable can be created by estimating (using only baseline covariate values) a propensity score for the household's uptake probability. Alternatively, one could estimate (again, using only baseline covariate values) the amount of treatment a household seeks based on the household's observable characteristics. Whichever one is chosen, it is then used for $\widehat{H}$ in Equation (6).

While exposure to a VIP latrine or safe drinking water is likely to be adopted by all households, the same cannot be said for PHAST training. Some households may attend while others do not. This can present a challenge for internal validity called self-selection into/out of treatment, leading to spurious comparisons and potential bias. This could occur if there are unobservable differences between the households that attend training sessions and those that do not. For example, those familiar with the importance of hygiene may choose to skip the training while those unfamiliar may not. This might mean that more training becomes associated with *poorer* hygiene (or at least reduces its average treatment effect).

Fortunately, this issue was foreseen and data collection requirements were put into place to disentangle and address direction of causality in the analyses. First, a set of questions in the IEMS will provide us with covariates correlated to the initial level of household hygiene knowledge and likely demand for training. Such variables can be used as either instruments or controls for baseline conditions. Second, NORC will have administrative data on the trainings by village. Together these data allow us to compute the average treatment-of-the-treated (ToT) effect as well the average intent-to-treat effect. Regardless, the present type of selection bias is likely only to underestimate effect size, not to overestimate it. Thus, if an effect is detected one can be confident that it is not due to overestimation.

[45] There is an emerging literature on the power of using propensity score methods within the context of CTV models. A good example of this literature is Hirano and Imbens (2004), "The Propensity Score with Continuous Treatments" in *Missing Data and Bayesian Methods in Practice: Contributions by Donald Rubin's Statistical Family*, New York: Wiley Press.

[46] In theory, this level also comprises two levels, the household itself and economically active individuals within it. However, for the present exposition, we ignore this complication.

4.3.4 Post-treatment revisions to the evaluation design

As is by now clear, the delay in construction of over 20 percent (11 villages) of treatment-group villages complicates the evaluation analysis considerably, compared to the evaluation design originally submitted. This subsection summarizes the implications, namely, why a revision to the original evaluation design is necessary and then discusses the pros and cons of administering the endline as originally envisioned in the data collection plan.

First, the facts. Construction delays prevented a subgroup of villages (Phase A_1 or equivalently, T_2 in Figure 1) in the original treatment group (Phase-A) — those in remote mountainous areas — from receiving the intervention according to the pipeline plan, which was the basis of NORC's original design (the right-hand-side of Figure 1).[47] This is clearly seen in Table 4.

Table 4: End of Construction Dates				
Phase	Group	Number of villages	Construction start-up	End of physical construction (water systems operational)
A^{rev}	Treatment	39	Dec 2010 – Feb 2011	June 2011 – Sep 2013
A_1	Treatment	11	Jan-Mar 2011	Jul 2013 – Feb 2014
C	Control	48	Jan-Apr 2013	Mar – Dec 2013

Source: Table 1 in Section 2.1.3 of this report.
Note: For reference, the IEMS Baseline was administered in December 2010 and the IEMS Midline in Nov-December 2012, with supplementary phase in April 2013.

The Need for Design Revisions

As a result of the construction delays, the original evaluation design became compromised. First, as explained in Section 4.3.2, the difference-in-differences estimator now must aggregate households whose length exposure to treatment varied over a much longer period than originally anticipated. The resulting average treatment effects, therefore, will be rather blunt and have a distribution with a larger standard deviation than one would like. Second (and also detailed in Section 4.3.2), MCC added several hypotheses (e.g., 1, 5, 8 in Table 3) after the original design was developed that, due to the way the intervention was designed, do not submit well to a difference-in-differences approach..

Third, the size of the treatment group available for use in the Period-1 evaluation, which was designed to cover anticipated short- and intermediate-run outcomes, was significantly reduced. (As shown Figure 1, T_2 villages had not been treated by midline). Hence the statistical power of the original design was reduced. The original evaluation of the joint effect was expected to achieve an MDES of between 0.23 and 0.28[48] and 80-percent power.[49] Without T_2 the experiment would likely only achieve 80-percent statistical power with an MDES of between

[47] While not listed here, there were several other compromising deviations in treatment implementation not foreseen in 2010 that nonetheless could also be addressed in the evaluation design revision.

[48] MDES is measured in standard-deviation units so these decimal values simply refer to the number (or proportion) of a standard deviation.

[49] This statistical power assumed an ICC of between 0.1 and 0.2, a level of significance of 5 percent, and covariate capture of 0.3.

0.27 and 0.32.[50] This may not sound like much of a loss, but for weak effects the extra power may make the difference between detection and no detection.

Fourth, the loss of T_2 villages probably compromised the representativeness of the original Phase-A treatment group, which was *non-randomly* split into subgroups, T_1 and T_2 (a concern we propose to test). If this split did result in T_1 and T_2 not being statistical similar, there are two consequences. First, one or more of the three quasi-experimental approaches described in Section 4.3.1 would be required to construct a valid control of villages from C_C to use for the baseline-midline impact evaluation. Second (and if an endline were not conducted), the external validity of the evaluation would be limited to households in villages like T_1 and not to those in T_2.

Is There a Need for an Endline?

With this assessment in mind, why, if we have a baseline and midline for some treatment and control villages, don't we simply use them for the evaluation instead of incurring additional costs (the endline) and complications to increase the sample size? What more would an endline offer? To answer this question it is helpful to reiterate the main objectives of the evaluation and then consider which, if any of them, would not be feasible without an endline.

NORC was asked to evaluate an overlapping set of treatments comprising:

- PHAST (training);
- Water system installation and maintenance;
- Setup, training, operation, and oversight of VWHCs and Water Minders; and
- Installation and maintenance of VIPs in households.

To what extent, then can the *revised* evaluation design achieve these objectives? First, the water systems with PHAST *can* be evaluated using 39 treatment villages and 50 control villages. Second, neither the effect of the water system without PHAST nor PHAST without the water system can be evaluated. This is because PHAST was implemented considerably *before* the IEMS baseline. Third, since no survey has yet been conducted on the Water Committees (VHWCs) or Water Minders (WMs), the effect of these two intervention activities cannot be determined unless the "AMP" surveys are conducted at endline and CTV models are estimated as per the revised design. Finally, the VIPs *can* be experimentally evaluated without an endline survey, but at a lower level of statistical precision than originally targeted. The original evaluation was expected to achieve an MDES of between 0.23 and 0.28 and 80-percent power while without T_2 the MDES may decrease to between 0.28 and 0.33.[51]

Returning to the original question, the addition of an endline, say in early 2015, would:

a. Provide an additional 11 treatment villages to the CTV, which would enable NORC to achieve higher levels of statistical precision and therefore greater credibility of its findings for the research questions.

[50] This estimated loss of detection, however, may still be an underestimate since it does not take into account the fact that 2 treatment villages would have been exposed to less than six months of treatment by the midline.

[51] Again, for this calculation we have assumed an ICC between 0.1 and 0.2, a level of significance of 5 percent, and covariate capture of 0.3.

b. Permit the Water-Committee and Water-Minders activities to be properly evaluated (via AMP surveys), since the effect of these actors needs to be assessed in light of contemporaneous household water use and perceptions (from an IEMS endline) and not in light of those from two years earlier at midline. At the same time, without the AMP surveys it will be hard to disentangle whether the degree of household water use (a key outcome) was due to (say) inadequate PHAST training and household idiosyncrasies or whether it was due to lapses in maintenance and lack of achieving targeted levels of water services.[52] This is also very important since MCC will want to know whether its community capacity-building (CCB) intervention activities were effective. While we do have (administrative) data on the training given to the WMs and VHWCs, (an "input"), only the AMP will give us definitive information on its resultant outcome.

c. Allow the rigorous evaluation of the long-term impacts of the WASH interventions on household productivity, income, and wealth – something MCC has long desired to measure but has not been able to in most compacts since by the time its activity components generate outcomes there is often too little time remaining in the compact to evaluate them. With no endline, the long-term effects of the MCC intervention on these key indicators of wellbeing are less likely to be detected with satisfactory power.

d. Improve the statistical precision of intermediate impacts estimated by the CTV models (since additional time will have passed for households requiring more than nine months to manifest impacts to have done so); likewise, since 50, rather than 37 treatment villages could be included in the sample, increases in statistical precision should be possible even for those households that manifest most of the improvement within nine months.

e. Introduce additional opportunities for evaluating the WASH intervention, such as inferring the evolution (trajectory curvature) of effects, i.e., are they increasing at a constant, increasing, or decreasing rate? This is of significant policy importance since it indicates whether most of the benefits come in the shorter term or whether they only arrive in the longer term.

f. Create opportunities to evaluate the research questions with multiple techniques using the same underlying data. This added level of robustness (which takes advantage of the unplanned splitting of both treatment and control groups) will increase the credibility of findings.

g. Expand the external validity of the evaluation to what was originally intended by also including households in villages like T_2 *as well as* those in T_1.

4.4 Limitations and Challenges to the Evaluation Design

Despite the thoroughness with which the evaluation design was conceived and executed, events in-country as well as issues related to the implementation of the intervention itself inevitably leads to threats to the validity of the analysis (e.g., randomization non-compliance and deviation

[52] Recall that IEMS endline and the AMP surveys are logistically connected: The HALEH module of the AMP surveys would be embedded into the IEMS endline and the WM and VHWC surveys are only useful when analyzed with contemporaneous (i.e., endline) IEMS data.

from treatment assignment status). This section points out the strategies we developed to minimize such threats.

Incidence of diarrhoea. Assuming that on the average a child under the age of 5 years has 2.2 attacks of diarrhoea per year, if data on diarrhoea are based on a 48-hour recall period, the frequency of positive answers to the question "has your child had an attack of diarrhoea that started in the past 48 hours?" will be 1.2%.[53] There is the possibility that a poor measure of diarrhoea (for example, one-month recall), will usually underestimate the beneficial effect of using an improved supply of water. The recall period in the Water Module of the IEMS is two weeks, which is the standard recall period applied in WHO surveys. Incomplete capture of the incidence of diarrhoea or field surveys during a dry season could also lead to such underestimation. Based on the experience at baseline, mitigation efforts were adopted for the IEMS midline and focused on giving enumerators specific training to give close attention to capturing this data accurately.

Contamination. The strength of randomly assigning villages to either treatment or control, is that the average values of any given characteristic of the two groups converge to become statistically identical as group size increases. This is important because the validity of the evaluation design depends on the control group serving as a good counterfactual of what the treatment group would have looked like at endline in the absence of treatment. The challenge here is that while geographically close treatment and control pairs increase the likelihood that they are mutually similar, it also increases the chance that proximity will result in the treatment village somehow contaminating the control village. For example, households in the control village might take advantage the new infrastructure in the treatment village. Such tensions can be reduced in a quasi-experiment through judicious choice of matching village pairs. However, in a pure experiment with truly random assignment of village to treatment, this is not as easy to prevent. This is especially true in the current case where NORC did not have full control over when treatment was dispensed.

There appear to be a half-dozen control villages that are potentially closer to treated villages than one would ideally like to have been the case (see the map in Annex B). However, several countervailing factors can be mentioned. First, the GIS coordinates used for the map in Annex B, though acquired from DRWS, did not appear to be fully accurate and so proximities may be incorrect. NORC endeavoured to use other sources to improve the coordinates but some distances may still remain larger than shown. Second, most villages were not treated at the start of the treatment period so the actual time available for contamination was shorter than the interval between the start of treatment of Phase-A villages and the start of treatment of Phase-C villages (the original control group). Third, the nature of treatment may have made it hard for one village to contaminate another. It is unlikely, for example, that households in one village would go to another village, no matter how close, to use a VIP. Learning from training, on the other hand, is easily conveyed, though in the present case all villages, regardless of when infrastructure construction was carried out, received PHAST training prior to the NORC baseline. Thus, such training *per se* could not be a source of contamination.

[53] Snyder, J. D., and M. H. Merson. 1982. "The Magnitude of the Global Problem of Acute Diarrhoeal Disease: A Review of Active Surveillance Data." *Bulletin of the World Health Organization* 60: 604–13.

Taking these features into account, we believe that contamination should not pose any first-order threats to internal validity. Still, NORC plans to test whether the effects detected are smaller the closer control villages are located to a treatment village.

Spillovers. Closely related to the issue of contamination is that of within-village spillovers. For example, how does one consider community effects when analyzing combinations of treatments – one household may have a VIP latrine that many other households might use; one household may get trained and then inform neighboring households of training.[54] As already indicated, in principle, all households in a treatment village got a VIP latrine so the only potential would be for "transitory" spillovers – where one household uses another's latrine as it waits for its own to be installed. All this could affect is the length of exposure to treatment if we had the exact installation dates, which we don't.

Validation of statistical methods. At this juncture, while baseline and midline household data have been collected, it has not been statistically confirmed that Phase-C (Figure 1) villages provide an adequate common support as a control group for Phase-A villages.[55] Likewise, the adequacy of econometric instruments to overcome selection bias and endogeneity in the CTV approach has not been statistically confirmed. Given the richness of the IEMS data, however, we are confident that an adequate instrumental variable can be identified or a propensity score can be estimated.

Regarding the challenges posed by selection bias to the RCT itself, we believe that the three independent techniques that we described provide a robustness of analysis and thus minimize the threat to internal validity, although perhaps at the cost of "degrading" the RCT to quasi-experimental methods.

Data from DRWS/Cowater Activity Monitoring Questionnaires. DRWS/Cowater developed and administered numerous questionnaires (the AMP questionnaires) to assess the construction of water systems as well as the quality and sustainability effects of PHAST training, latrines, and VWHC activities.

In 2010, Cowater supported DRWS in developing a comprehensive plan to monitor all activities associated with the construction of the compact rural water systems, which is described in detail in the July 2010 *Department of Rural Water Supply & PMCS – Cowater International, Activity Monitoring Plan* (updated in March 2012). The Activity Monitoring Plan (AMP) was intended to assist the DRWS in "keeping track of progress in project implementation and the various project components in relation to targets, timelines." Towards this end, Cowater, in partnership with DRWS and MCA, developed a series of one-page checklists/data collection sheets to record progress on water supply-related outputs that measured progress on construction of water supply systems; compliance to environmental, health and safety requirements associated with construction; functionality of Village Water and Hygiene Committees; hygiene awareness of villagers; latrine hygiene; and knowledge and training levels of Water Minders.

[54] We thank an MCC reviewer for posing this question.

[55] MCC specifically requested that NORC hold off on conducting any data analysis until they have approved the present document.

As per the AMP the DRWS district staff were to collect data at the start of construction, end of construction and on a monthly basis during construction. The submitted data was to be inputted into the AMP database and then included in the Cowater monthly reports. However, due to several challenges - lack of internet at the district level, long distances to travel for paper submission – district staff were hampered in their attempts to submit data from the field. Therefore, there are major gaps in the availability of actual AMP data collected in the field. Cowater monthly progress reports progress and ten audit reports conducted by Cowater covering a sample of villages report on many indicators in the AMP; however, the source of this data is not the AMP database, but progress reports from Cowater Site Engineers' site visits, and discussions with contractors and DRWS districts.

NORC's primary concerns about the existing AMP data (as recorded in monthly reports) as an input into the impact evaluation are three-fold: (1) while they comprehensively report on tangible measures such as completion dates of construction activities and trainings, information on indicators related to sustainability, such as knowledge and competency of VWHCs and Water Minders, or hygiene awareness of communities is not reported on as consistently across time and villages; (2) data on hygiene-awareness and functionality/competence of Water Minders and VWHCs are not available for all Phase A and Phase C sample villages; and (3) this data will not be available for the villages, post compact, which limits the ability to measure sustainability beyond the implementation period.

In light of these concerns, NORC refined and expanded the AMP questionnaires that assess hygiene and sanitation awareness, knowledge, and practices; the functionality of VWHCs in providing maintenance and operational support to village water systems; and the capabilities and functions of the Water Minders, making them more consistent with evaluation needs. Should MCC authorize an endline IEMS survey, NORC proposes to take advantage of the now-completed AMP instruments by introducing it as a module in the endline survey (based on recall questions). See Section 5.2 of this report for further details.

Data quality of IEMS. NORC conducted a thorough data quality review of each round of data collection and, with the exception of two data-collection mishaps by the Bureau of Statistics (BoS) – one in the baseline and one in the midline – found the quality to be satisfactory. These exceptions are described in detail in Section 5.1. Concerning the baseline, there were delays by BoS in revisiting the field to rectify their improper execution of disposition coding, which may have implications for bias in variables of interest. Concerning the midline, for unexplained reasons, in 75 villages BoS ignored the fact that they were collecting panel data and interviewed *new* households instead of returning to the same households as for baseline. As a result, BoS had to return to the field later to interview the missing baseline households. This fragmentation of the midline data collection poses threats to the evaluation design and may threaten its internal validity. As discussed further below, at this juncture, we do not know whether this has threatened internal validity of the design[56].

[56] Analysis of these data was put on hold by MCC, pending first, the documentation of a program logic and then, the approval of a revised design report that meshed with the documented program logic.

5. DATA SOURCES AND DATA COLLECTION

The main data sources for this evaluation are the Impact Evaluation Multipurpose Survey (IEMS) and the suite of Activity Monitoring Plan (AMP) surveys[57].

5.1 Impact Evaluation Multipurpose Survey (IEMS)

The IEMS is a longitudinal analytic survey specifically designed to collect data for the impact evaluations of the MCA-Lesotho Compact health and water activities. Three rounds of the IEMS were originally proposed in the evaluation design: a baseline, midline, and endline. These three rounds were approved by MCC. To date, the Bureau of Labor Statistics (BoS), under an Implementing Entity Agreement (IEA) with MCA Lesotho, has conducted the baseline and midline IEMS.

The baseline IEMS was conducted in December 2010, prior to the start of the construction of water and hygiene systems between December 2010 and March 2011 in treatment villages. As is evident from Table 1, however, PHAST training in the vast majority of Phase A and some Phase C villages preceded the baseline data collection. Therefore, the December 2010 data collection only serves as a true pre-intervention baseline for the construction of water and hygiene systems. The baseline data collection covered treatment and control (Phase A and Phase C) villages for the rural water intervention. (It also covered villages and enumeration areas for the urban water and health sector activities.)

As mentioned in Section 4.4, data collection by BoS suffered mishaps in both baseline and midline data collections. NORC submitted a data quality review of the baseline data collection to the MCA for rectification by the Bureau of Statistics (BoS). While BoS finally complied, the time elapsed made it difficult for BoS to adequately revisit the field. One consequence was that disposition coding was not properly executed. It is not clear whether or what bias this might imply for each of the variables of interest. More generally, the quality of the dataset will have implications for the degree of statistical precision of the evaluation's statements of attribution.

In November-December 2012, BoS conducted a midline follow-up data collection. The intent was to collect panel data from the sample of households from the baseline. Towards this end, NORC developed and provided BOS with detailed sample lists of households to visit in each treatment and control village. Upon receiving the complete dataset from BoS and conducting a data quality review, NORC discovered that, while BoS visited all the requisite villages and interviewed the requisite 13 or more households per village, only a subset of these households were from the list NORC provided of households in the panel study. In as many as 75 villages, enumerators had unilaterally chosen not to revisit 312 baseline households, and had instead interviewed new households to arrive at the requisite numbers. As a result of this omission, NORC had BoS return to the field in April 2013 to complete correctly the midline data collection for the

[57] In the early stages of the evaluation, NORC and MCC agreed to use the ongoing Continuous Multipurpose Survey (CMS), conducted on a quarterly basis by the BoS as the vehicle through which to collect data for the three MCC Lesotho evaluations. As such, a health and water module were added to the CMS questionnaire and fielded as a supplement to 2 rounds of the CMS. However, later, to ensure a more comprehensive coverage of treatment and control villages and EAs, the CMS data collection was replaced with the IEMS. We do not plan to use CMS data for the impact analysis; however, it may prove useful, given it national representativeness, to check for external validity of the evaluation design.

missing baseline households. This fragmentation of the midline data collection poses threats to the evaluation design. For example, it means that in control villages, midline data collection in some households occurred after construction of water systems had already commenced. At this juncture, we do not know whether or not the returns were random across our full sample or more predominant in treatment or control villages since baseline analysis has been delayed (see footnote 56).

This fragmentation of midline data collection may pose threats to the internal validity of the evaluation design. First, the delay in administering the midline to those control villages meant that midline data collection of some households occurred *after* construction (but probably not operation) of their water systems had already commenced. Second, (i) if the reasons that BoS enumerators initially skipped and had to return to a panel household were somehow correlated to factors modulating intervention impact and (ii) if the initially skipped households were not random across our full sample, then this could introduce bias into the results of some RCT applications (though not the CTV models).

The final IEMS was originally planned for November-December 2013. At this juncture, however, whether MCC will actually fund an endline data collection is undecided.

Sample Frame

The sampling frame for the IEMS consists of all villages in Lesotho based on publicly available geospatial data and 2006 Census data. Information on administrative location, geo-coordinates, rural-versus urban designation and population was merged with publicly available physiographic and geographic data to be used as covariates in the sampling. From this central dataset, individual sample frames were designed and PSUs were selected for two central project components: rural water and urban/peri-urban water. For rural water, villages were the primary sampling units.

The centralized frame dataset consists of the following variables:

- Primary Sampling Unit identifying information: Village name, Village ID (both GIS ID and Census ID), Village Geo-coordinates (X and Y), Enumeration Area ID, Community Council, District, Constituency
- General covariates: Population, average annual temperature, precipitation, vegetation productivity potential, number of households, urban/rural designation

Sampling Design

The random sample selection was sequentially sampled without replacement in the form of a multistage cluster design as follows:

There are two stages to the design for the rural water intervention. They cover the designation and selection of villages (PSUs, clusters) and households (SSUs):

Village sample. As described in Section 2.1.3, of the 250 villages in 10 districts selected by DRWS for the MCA rural water interventions, 100 villages (10 per district) were deemed "ready" for the intervention in 2008. Fifty of these 100 villages were randomly assigned to treatment (Phase-A), while the remaining 50 were assigned to the control group (Phase-C). Final implementation lists, however, only consisted of a random sample of 50 treatment villages, 48 control villages[58]. The village locations are shown on the map in Annex B. The only difference between the treatment and control villages was that the control villages would receive treatment after a delay during the evaluation period according to the pipeline design

Unfortunately, infrastructure construction ran into delays in villages selected for treatment in three districts (Mohkohtlong, Qacha's Nek, and Thaba Tseka). These districts are the most remote (and mountainous), hence, likely to be systematically different from the other treated villages. The emergence of this group receiving later treatment required our evaluation design to be retrospectively divided into two treatment groups, T_1 (earlier treatment) and T_2 (later treatment).[59]

Household sample. Within each treatment and control village a systematic random sample of 13 households was selected[60]. The interview was conducted with the head of the household or the person in the household most knowledgeable about household water and sanitation issues.

5.2 Activity Monitoring Plan (AMP) Surveys

As described in section 4.4., the AMP surveys are a suite of checklists and data collection forms developed by Cowater, in conjunction with DRWS and MCA, to monitor the construction and training activities associated with the Lesotho WSP. Of these, NORC has refined and expanded three surveys to collect data needed for the impact evaluation. These surveys would be administered to community (household) members, VWHC members and Water Minders in rural water villages along with the endline IEMS. The sample for these surveys would overlap with the villages and households in the IEMS.

The purpose of the surveys, which we propose to implement with the endline IEMS, are described below:

- *Hygiene Awareness & Latrine and Environmental Hygiene* (HALEH). Assesses the household's knowledge and practices of proper hygiene and sanitation. The questionnaire consists of two parts: PHAST awareness and observation of the VIP latrine and the household's environmental surroundings. The HALEH sample will consist of 13 interviews per village, administered to households that were interviewed during the IEMS baseline and midline. For efficiency HALEH will become an additional module of the IEMS questionnaire at endline.

[58] Note that the 5 rural villages could not be directly matched with data in the master sampling frame.

[59] Removing T2 from the original treatment group of T1+T2 affects the external validity of T1: it cannot be used to infer responses of T2 to treatment.

[60] Prior to conducting the IEMS, BoS had conducted a listing of all of their Enumeration Areas. Each EA consists of several of villages. The listing of households within an EA starts from the outer edge of this cluster of villages at a recognizable structure such as a church, store, or health facility identified by BoS' GIS team. The northernmost household to that structure is listed as Household 1 within the EA area. Then all other households are numbered in order in a clockwise direction starting from the outer circumference and moving inward in a circular fashion throughout the entire EA. The final stopping point of the listing is the last household at the very center of the EA.

IEMS required the sampling of villages. To sample from each villages, BoS organized each village's household list in numerical order in excel. From each of the village household lists, BoS utilized Excel's RAND function to select a random starting point at which to begin systematically sampling from the village lists. An appropriate sample interval was selected according to the number of households within a village and systematic sampling was carried out to obtain the required sample size for each of the IEMS villages.

The HALEH module will provide an endline assessment on the link between the level of hygiene training received by the beneficiary and hygiene and sanitation knowledge and practices.

- *Village Water and Health Committee Functionality* (VWHC). Assesses the degree to which the members of the VWHC are able to provide monthly routine operational management to the village's water system. It also assesses the current status of the system's operation. The VWHC questionnaire will be administered to the committee in a group setting at endline.
- *Water Minder Expertise* (WM). Assesses the level of training that the WM received by determining his level of knowledge in after care maintenance and in repairs of the system. It will be administered to each Water Minder in the village (approximately 2-3) at endline.

We expect the PHAST training and aftercare training conducted by DRWS for communities to differentially impact indicators of interest. For example, the training of VWHCs and Water Minders is expected to affect the sustainability and reliability of water systems. A key policy document on water mentions that at least 20 percent of water systems are not functioning at any given point in time and other official documents have a higher figure.[61] Intermittent supply will weaken the treatment. The capability and functionality of the VWHCs and Water Minders, measured using data collected through the two related AMP questionnaires, will likely affect the continuity of supply and, therefore, could be an important factor in accounting for differential impact. Where supply reliability is found to be significant over the treatment period, NORC will incorporate an indicator of reliability into its analysis.

Similarly, although all households in all villages had access to PHAST community training since before the start of construction, varying intensity of the PHAST training and level of household participation in these trainings, measured through questions in the HALEH questionnaire, is expected to yield varying degrees of hygiene awareness, which in turn is likely differentially impact higher level indicators such as incidence of water-related disease.

The AMP questionnaires were initially developed by DRWS as simple checklists; however, with the impact evaluation and measurement of sustainability in mind, NORC fleshed out these checklists into more complex questionnaires for DRWS to administer through their CLOs. Should MCC authorize an endline household (IEMS) survey, NORC proposes to take advantage of the now-completed AMP instruments by introducing it as a supplement to the endline IEMS.

Sampling Design

The sampling design for the HALEH survey is identical to that of the IEMS since the former would simply be a module in the latter. All WMs in a village (approximately 2-3) will be interviewed. All members of the VWHC in each village will be interviewed in a group setting.

[61] Lesotho: Water running on empty. www.irinnews.org/Report/.../LESOTHO-Water-running-on-empty. See also Interim Strategy for the Water and Sanitation Sector, Ministry of Natural Resources, Section 2.1. The DRWS After-care Strategy reports that DRWS estimates that 25% are not functioning fully and delivering the design standard of 30 liters/person/day (p1).

These surveys will be administered by the same enumerators administering the endline IEMS, with the HALEH being administered as a module of the IEMS questionnaire, and the other two surveys being administered separately, but during the same visit to a village.

In our opinion, it makes little sense to conduct the AMP without also conducting the endline IEMS. As shown in Table 1, the data gathered through the AMP surveys serve as treatment indicators that are directly linked to the outcome indicators collected through the IEMS. Linking AMP data collected in 2014 or 2015 to 2012 IEMS data is unlikely to serve as the foundation for a credible impact evaluation. It is possible to ask respondents about their sanitation and hygiene practices in 2012, but recall errors are likely to be very high. As such, our recommendation is that MCC choose to conduct both the IEMS and AMP simultaneously or abandon both data collection efforts[62].

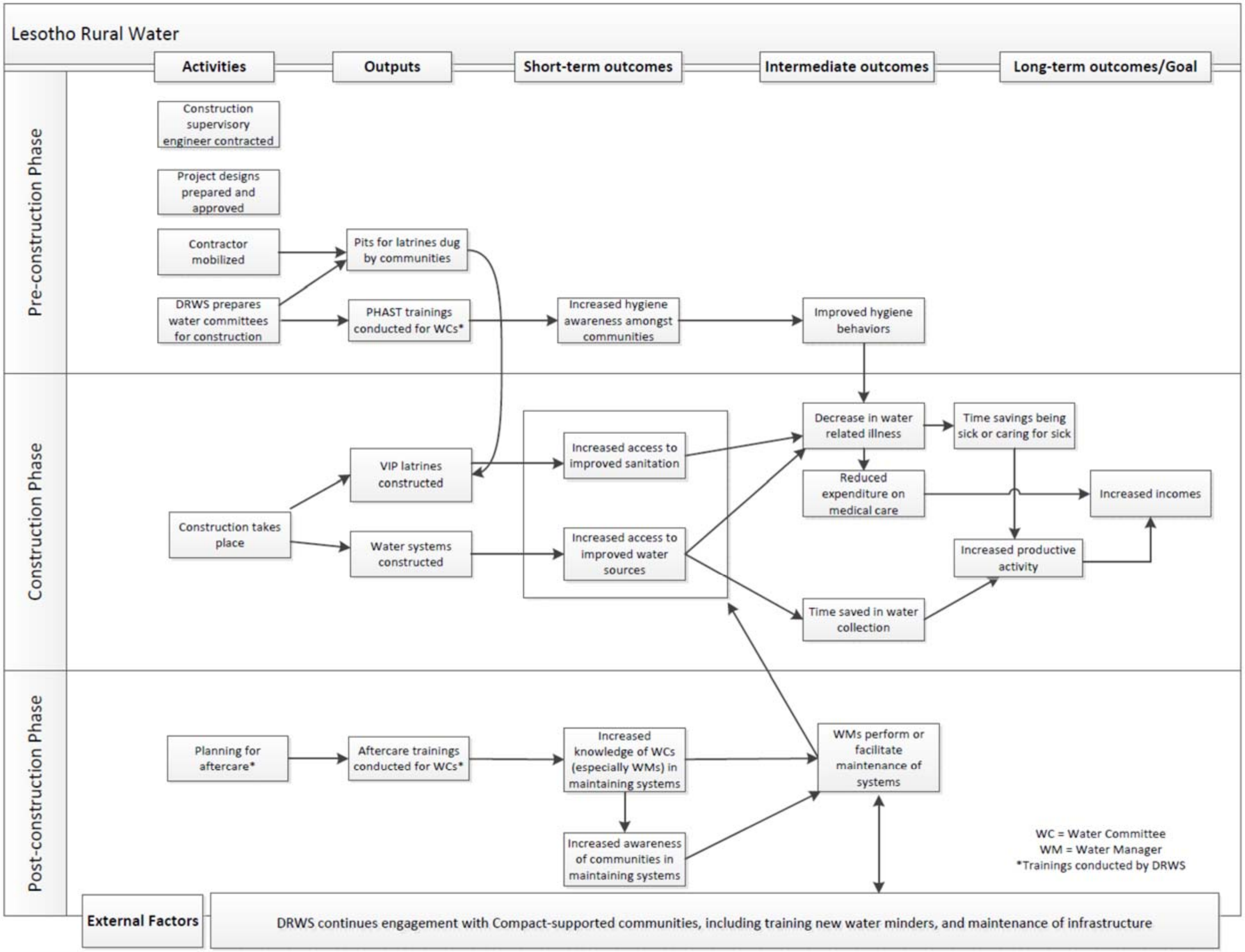
5.3 Focus Groups

Focus group discussions (FGDs) in rural water villages, if conducted, will have three functions. First, they will help to put into context the findings, whether positive or negative, from the statistical analysis of the quantitative data. Second, they will offer a way to better interpret statistical results that were not statistically significant. Third, they will be used to complement the quantitative analysis described above by allowing the assessment of perceptions of topics that are not conducive to quantitative inquiry.

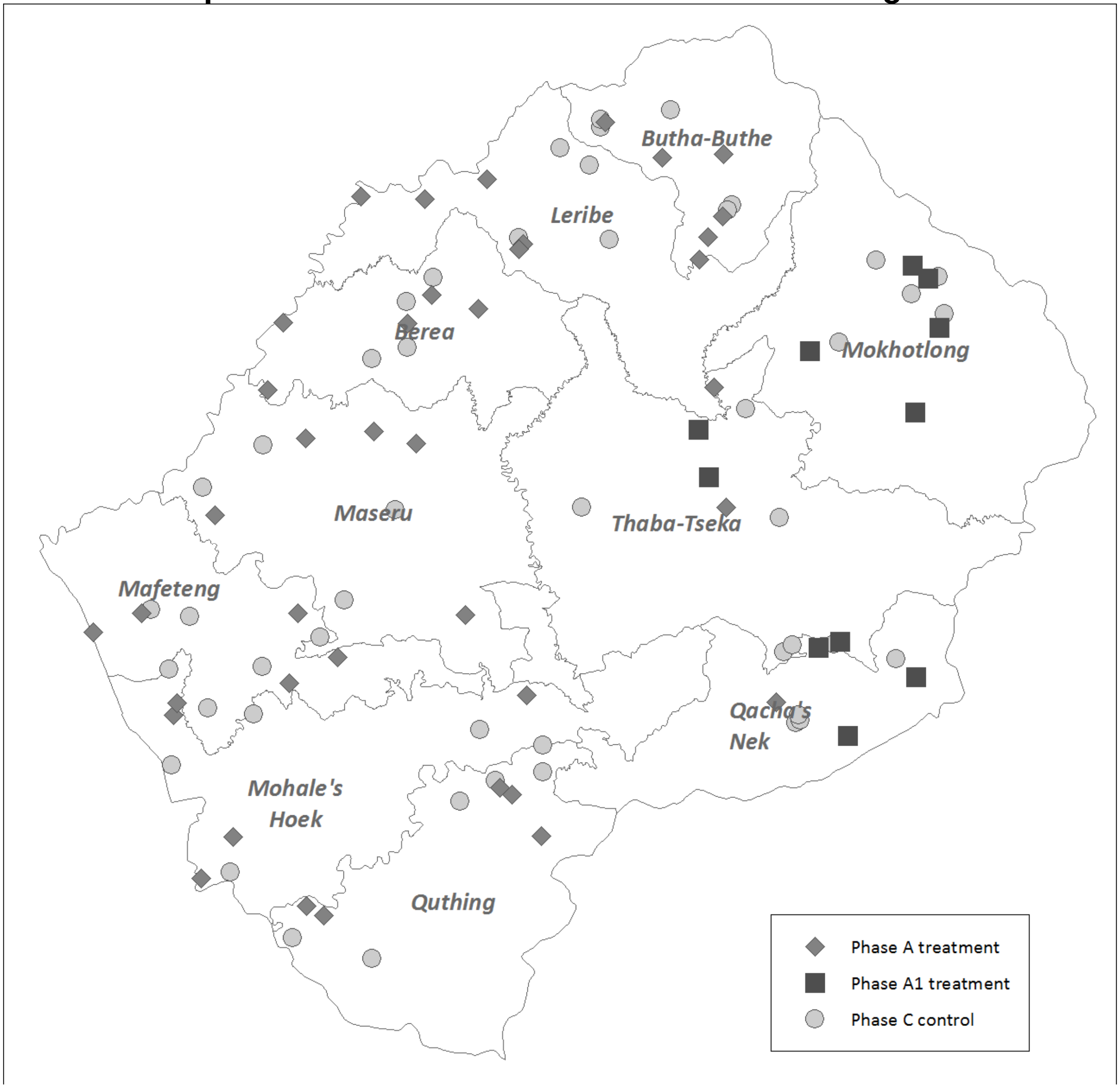
We will determine the number, composition, and content of FGDs in conjunction with MCC.

[62] A MCC reviewer requested that NORC conduct power calculations and estimate sample sizes for the AMP, in the event that the AMP surveys are conducted without the endline IEMS. As discussed in the main body of the report, we do not believe that there is much point to doing the AMP surveys by themselves. As such, we have not estimated these sample sizes.

Annex A: Program Logic of Lesotho Rural WSP (Updated 2014)



Annex B: Map of locations of treatment and control villages



Note: Some of the geographic coordinates provided by DRWS were inaccurate or insufficiently precise. NORC staff attempted to verify the location of each rural water village using the Lesotho 2006 Census GIS database and external map sources (e.g., Google Earth). While we are highly confident about the locations of the majority of sites, some of the points shown on the map may not represent the exact location of treatment and control sites.

Annex C: Review of Literature Pertinent to the Rural Water Evaluation

Author and title	Intervention	Methodology & data	Relevant conclusions
Effects of increased use of safe drinking water, use of a VIP and better hygiene behavior on incidence of water-related diseases (and other health outcomes).			
WASH (Water, Sanitation, and Hygiene) and Water/Sanitation (no hygiene) Interventions			
Huttly et. al. (1990). The Imo State (Nigeria) Drinking Water Supply and Sanitation Project	Water Supply and Sanitation Project in 3 villages in south-eastern Nigeria	Difference in difference; repeated cross-sectional surveys and longitudinal data ; variables were water diseases (worms and diarrhea, all ages) and nutritional status (children)	No impact on overall period or prevalence of disease in cross sectional data but impact on diarrhea morbidity was found in limited sub-groups. Also significant overall program impact was found on worm disease in longitudinal data. A greater impact of water availability rather than quality was suggested for rates in young children. The prevalence of wasting among children < 3 years decreased significantly with intervention.
Akuoko-Asibey and McPherson (1994). Assessing hygiene and health related improvements of a rural water supply and sanitation programme in northern Ghana	Rural water supply and sanitation programme in northern Ghana	Quasi experimental difference in difference; variables were sources of water collection and storage arrangements, waste disposal, hand washing, food storage arrangements, and defecation and practices associated with feces in program and non-program areas	Some success in water use and water and food storage; little progress in attitudes towards disposal of HH and human waste and in women's knowledge of disease.
Children-specific WASH Interventions			
Aziz *et. al.* (1990). Reduction in diarrheal diseases in children in rural Bangladesh by environmental and behavioral modifications	Water, sanitation and hygiene education intervention in a rural area of Bangladesh	Difference in differences estimator between two similar areas	Children in the program area had 25% fewer episodes of diarrhea than those in the control area. Within the program area, children from households living closer to hand-pumps or where better sanitation habits were practiced experienced lower rates of diarrhea. Suggests that an integrated approach to interventions can have a significant impact on diarrheal morbidity.
Checkley *et. al.* (2004). Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community.	No intervention *per se*; studies a birth cohort of Peruvian children, for health outcomes over 24 months, against baseline data on household water and sanitation conditions.	Difference in difference. Survey data followed up with children once a day for diarrhea and once a month for anthropometry; obtained baseline data on household WASH conditions.	Better water source alone did not accomplish full health benefits. In 24-month-old children from households with a water connection, those without adequate sewage disposal and with small storage containers were 1·8 cm shorter than children in households with sewage and with large storage containers. Children with the worst conditions for water source, storage, and sanitation had 54% more diarrheal episodes than did those with the best conditions.

Author and title	Intervention	Methodology & data	Relevant conclusions
Nanan et. al. (2003). Evaluation of a water, sanitation, and hygiene education intervention on diarrhea in northern Pakistan.	WASEP, a village level project which incorporates engineering solutions with appropriate education to maximize facility usage and improve hygiene practices.	Difference in difference (3 months only), using logistic regression and controls.	Children in control villages had 33% higher odds of diarrhea than children living in program villages. Boys had 25% lower odds of having diarrhea than girls. Relevance: shows how integrated approaches can be successful.
Esrey *et. al.* (2000). Drinking water source, diarrheal morbidity, and child growth in villages with both traditional and improved water supplies in rural Lesotho	Villages in four districts (50 per cent of the rural population of Lesotho) in the lowlands or foothills received an improved water supply between 1967 - 1983, in form of a continually functioning tap or hand pump.	Post-post multivariate regression. Data included morbidity and growth data on 247 children 5 and under in 10 representative villages that had an improved water supply at least one year prior to investigation. Compared those who relied exclusively on improved water to those who relied on it only partially.	Children whose families relied exclusively on the new water supply for their drinking and cooking needs grew 0.438 cm and 235 g more in six months than children whose families supplemented the new water supply with the use of contaminated traditional water for drinking and cooking. Results suggest that improved drinking water supplies can benefit preschool children's health after infancy, but only if they are functioning and utilized exclusively for drinking and cooking purposes
Hoque *et. al.* (1996). Sustainability of a water, sanitation and hygiene education project in rural Bangladesh: a 5-year follow-up	An integrated water supply, sanitation and hygiene education project from 1983-87. Provided hand pumps, pit latrines, and hygiene education to about 800 households. After 1987 no external support was provided to maintain these provisions.	Methodology: Difference in Difference using a cross-sectional data. Data: Baseline and follow-up surveys from 500 randomly selected households from the intervention and control areas.	82% of the pumps still functional 64% of latrines and 84% of the adults were using latrines (vs. 7% in the control area.) The prevalence of diarrheal diseases among the control population was about twice that among those in the intervention area. However, knowledge related to disease transmission, was poor (though, recall that all training had stopped) and similar in both areas. Relevance: importance of continuance of WASH education.
Water-Only Interventions			

Author and title	Intervention	Methodology & data	Relevant conclusions
White, Bradley, and White (1972). Drawers of Water: Domestic Water Use in East Africa	Improving water volume and quality	Using observational evidence only in 34 sites in 3 countries over 3 years.	Shows positive effect of water quality on water borne diseases; one of the original studies in the field and still considered by some the most comprehensive study to date
Kolb et. al. (2008). An integrated method for evaluating community-based safe water programs and an application in rural Mexico	UV Waterworks, a community-based water purification system in rural Mexico	Stepwise evaluation framework of effect on health 5 years after the programme began; variables include physical performance of the water system, community capacity to maintain and manage the systems, and the time and bud-get constraints of households participating in the program.	No impact on diarrhea incidence was found. (a) household priorities and preferences were a key factor in maintaining exposure to safe drinking water sources, and therefore (b) user convenience was a primary leverage point for programme improvement.
Hygiene-Only Interventions			
Wilson *et. al.* (1991). Hand-washing reduces diarrhea episodes: a study in Lombok, Indonesia	Sixty-five mothers were given soap and an explanation of the fecal-oral route of diarrhea transmission, and the message was repeated and rein-forced fortnightly	Pre-post evaluation of 65 families	Children of these mothers experienced an 89% reduction in diarrhea episodes
Independently of other interventions, the impact of VIPs on incidence of water related disease (children only)			
Semba *et. al.* (2011). Relationship of the Presence of a Household-Improved Latrine with Diarrhea and Un-der-Five Child Mortality in Indonesia	Improved latrines in rural and urban areas of Indonesia	Multivariable logistic regression models	The lack of a household improved latrine is associated with diarrhea and under-five child mortality in Indonesia.
Daniels *et. al.* (1990). A case-control study of the impact of improved sanitation on diarrhea morbidity in Lesotho	Project to improve sanitation in Mohale's Hoek district within Lesotho	Randomized case-control design using clinic-based diarrheal cases and controls that experienced other illnesses. Data collected on child, illness, access to water/sanitation, hygiene practices from interview with caregiver. A ran-dom sample visited at their homes and the water / sanitation facilities/general conditions were observed.	Children under-5 from households with a latrine experienced 24% fewer episodes of diarrhea then households without a latrine. The results of this study provide evidence that improved sanitation can have a positive impact in the reduction of diarrhea morbidity in young children in rural Lesotho.

Author and title	Intervention	Methodology & data	Relevant conclusions
Andres *et. al.* (2014). Sanitation and Externalities : Evidence from Early Childhood Health in Rural India	Household level: moving from open to fixed-point defecation OR from unimproved sanitation to improved sanitation. Village level: an external benefit (externality) produced by the neighborhood's access to sanitation infrastructure	Individual production function of health for children assuming a linear-in-parameters approximation, and robust results using several econometric specifications. Main inputs are HH access to sanitation and ratio of access to sanitation at village level. Data: 206,414 children under 48 months in rural areas of India from District Level Household Survey '07-'08.	Finds significant direct benefits and concave positive external effects for both improved sanitation and fixed-point defecation. At HH level, finds 10% reduction in diarrhea from sanitation improvements and 5% reduction from moving from open to fixed point defecation. Combining HH and village interventions finds 47% reduction in diarrhea prevalence between children living in a household *and* a village with improved sanitation vs children without either. 1/4 of benefit is due to the direct benefit rest to external gains.
Effect of improved access to safe drinking water and improved hygiene practices on time-savings to households (collection time, days sick, hours off work)			
Ilahi (2001). Gender and the allocation of adult time: evidence from the Peru LSMS panel data	None; uses panel data from Peru in 1994 and 1997 to examine the impact of water infrastructure on total time spent in housework and in income-generating activities by male and female adults.	Regression estimation using cross sectional data, controlling for unobserved heterogeneity. Data from 1994 and 1997 Peru LSMS panel household data: 898 HH's and 2095 individuals.	Women in households without in-house water supply do not have significantly higher housework burdens than women in households with piped nor do they spend less time in income-generating activities. For men, however, in-house water supply significantly increases time spent in self-employment activities (such as agriculture) and decreases time spent in wage work. Demonstrates potential positive income-generating impacts of piped water supply for men but questions it for women.
Ilahi *et. al.* (2000). Public Infrastructure and Private Costs: Water Supply and Time Allocation of Women in Rural Pakistan	None, uses existing HH surveys to examine access to water- community and household levels and the time allocation of women	Estimation reduced-form time equations. Data from the 1991 Pakistan Integrated Household Survey (2,400 rural household)	Poor water supply induces women to reduce their market-oriented work and thus their contribution to household income (however, impact on overall HH income unclear because male labor response is unclear). Results also indicate poor water supply causes an increase in the total work burden of women and a decrease in their leisure. Improved water supply could change the nature of women's contribution to the household from performing everyday chores to doing income-generating work.

Author and title	Intervention	Methodology & data	Relevant conclusions
Boone et. al. (2011). Household Water Supply Choice and Time Allocated to Water Collection: Evidence from Madagascar	None: uses household survey data from Madagascar	Conditional logit model for household's choice of water source as a function of distance to the source and HH characteristics. Reduced form model regressing collection time on a set of exogenous variables. Data: 2190 household surveys with detailed data on the characteristics of household members, as well as information on household wealth and assets. The employment and time use module included information on time use in various activities in the last seven days, including hours spent gathering water.	Women and girls spend the most time gathering water. However, investments to reduce to the distance to water sources will have larger impacts on adults than children, and on men than women.
Effects of improved access to safe drinking-water and improved hygiene practices on school absenteeism (sick days, school days missed)			
Nankhuni and Findeis (2004). Natural resource-collection work and children's schooling in Malawi.	None-uses existing data and hypothesis only that investigates if long hours of work spent by children in fuel wood and water-collection activities influence the likelihood that a child aged 6–14 attends school.	Two-stage conditional maximum likelihood estimation. Data from the 1997–1998 Malawi Integrated Household Survey (IHS) conducted by the Malawi National Statistics Office (NSO) in conjunction with the International Food Policy Research Institute.	Attendance decreases as hours allocated to resource collection work increase. Having piped water access in the home significantly reduces the probability of and time spent in water collection among children. They also find that having piped water access is positively associated with a child attending school and not doing any water-collection and negatively associated with water collection while attending school. Girls spend more hours on resource-collection work and are more likely to be attending school while burdened by this work and may find it difficult to progress well in school. However, girls are not necessarily less likely to be attending school (may do both collecting and attending). Shows water programs can affect absenteeism and that, furthermore, they can especially have an impact on girls' attendance.
Ilahi (2001). Children's Work and Schooling: Does Gender Matter? Evidence from the Peru LSMS Panel Data	None; uses data from the Peru LSMS Panel Data	Uses reduced-form equations for each individual, controlling for unobserved heterogeneity using the panel properties of data. Data from Peru's LSMS panel on time allocation of boys and girls in both rural and urban areas.	The findings suggest that changes in household water supply affect the schooling and work of girls more than boys. An in-house water supply has a significant impact on grade-for-age of girls but not boys, and seemingly no significant impact on time in housework.
Effects of improved access to safe drinking-water close to home on productivity of economically active members of household (agricultural output, income per hour, days worked)			
See Crow *et al.,* below.			

Author and title	Intervention	Methodology & data	Relevant conclusions
Kiendrebeogo (2012). Access to Improved Water Sources and Rural Productivity: Analytical Framework and Cross-country Evidence	None	Regresses rural labor productivity growth rate on access to drinking water (with controls). Based on sample of 27 African countries over 1990-2010.	The empirical analysis reveals that increasing the access rate to drinking water significantly increases the growth rate of agricultural labor productivity. Surprisingly, the access rate to improved sanitation facilities has not significantly impacted rural productivity growth. However, the positive effect of drinking water access is reinforced by the presence of a better sanitation system.
Effects of time savings gained through improved access to safe drinking-water and reduced incidence of water-related diseases on household income			
Crow *et. al.* (2012). Community Organized Household Water Increases Not Only Rural incomes, but Also Men's Work	None; took a census to compare results of Community-organized, household water improvements in western Kenya and divided villages into 3 arms: (i) unprotected; (ii) protected but not piped; and (iii) protected and piped.	Used mixed methods for (1) with/without comparisons between households obtaining water from different types of springs – unprotected, protected and not piped, and protected and piped; and (2) before/after comparisons for households with protected and piped water. The before/after data is based on respondents' recall. Quantitative survey for prevalence of different water management regimes, and qualitative for characteristics of communities. Data: A spring census was as a sampling frame for the selection of seven villages. Two villages have protected springs and piped homestead connections; two have protected springs but no homestead connection; and three draw water from unprotected springs.	Piped water reduces the work of women and girls, and facilitates home garden and livestock production. Women recognize clear time-benefits. Men, however, experience extra work. Together these changes lead to increased household incomes.