Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Amber Dismer Date

Rural Kenyan Household Stored Water Quality

By

Amber Dismer
Master of Public Health

Global Environmental Health

_____

Dr. Clair Null
Committee Chair

_____

Dr. Paige Tolbert
Committee Member

Rural Kenyan Household Stored Water Quality

By

Amber Dismer

Bachelor of Arts
Augustana College
2007

Thesis Committee Chair: Clair Null, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Environmental Health
2012

Abstract

Rural Kenyan Household Stored Water Quality
By Amber Dismer

Background: The Joint Monitoring Program recommended collecting presence absence tests of households' stored water quality after 2015 for all future demographic health surveys, multiple cluster indicator surveys, and living standards measurement surveys. The main objective of this thesis is to provide guidance on collecting clustered water quality measurements for practitioners and health scientists to quantify gains in households' and springs' water quality.

Methods: The Spring Improvement Project Household Survey's cluster randomized controlled trial design in western Kenya provides a unique opportunity to test the presence absence method against the logarithmic MPN of *E. coli*, the World Health Organization's drinking water risk-levels, and the geometric means method. Clustered analyses were performed for two sampling timeframes across months of the year and across bi-weekly rounds for households' stored water quality. Similar analyses were conducted for the springs where water was collected.

Results: Over 15 bi-weekly data collection rounds, households' stored water quality did not differ across the months of the year or the bi-weekly round. The logarithmic MPN of *E. coli* method ($F= 3.83$, $p=0.0001$) and the presence absence test (Wald $x^2= 27.88$, $p=0.03$) detected significant variability of springs' water quality by bi-weekly round. The intra-cluster correlation coefficients (ICCs) for households' stored water in spring clusters (logarithmic: 0.09; presence absence: 0.04) were smaller than the ICCs of clusters of springs' water across time (logarithmic: 0.46; presence absence: 0.10).

In multi-year rounds, monthly variability of households' water quality was detected at the baseline by all four methods. However, only the logarithmic MPN of *E. coli* method and the WHO risk levels detected baseline differences of springs' water quality by month.

Conclusion: The presence absence test yielded the same results for households' water quality as the logarithmic MPN of *E. coli* method; however, more samples per cluster are required for water quality interventions.

Rural Kenyan Household Stored Water Quality

By

Amber Dismer

B.A.
Augustana College
2007

Thesis Committee Chair: Clair Null, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Environmental Health
2012

Acknowledgements

My deepest thanks and admiration go to Dr. Clair Null for her guidance, patience, and assistance throughout this past year. Your constructive comments, questions, desire to dive into data, structure to answer research questions, and collaborative spirit has been inspiring. It has been both a pleasure and an honor to work with you. Thank you very much!

I am also very grateful to the following people. Dr. Karen Levy, thank you for providing essential assistance on analyzing and presenting water quality data at key turning points in the thesis! I sincerely appreciate the support and editing help from Cassie O'Lenick, Maggie Hodges, and Janine Dismer. Thank you all! I claim any remaining errors as my own.

There are many of you whose support and encouragement has also made this an enjoyable process. You have delivered chicken noodle soup through the busy times, called, baked with me, and offered new statistical ideas. We have also taken relaxing breaks together. Çox sag olun to all of my family, mentors, colleagues, and friends! I want to especially thank Flavia Traven, Danika Barry, Gaëlle Gourmelon, Marissa Grossman, Dana Brosig, Carrie Mc'Neil, Dr. Philip Brachman, the Humphrey Fellows, Heather Reese, Priyanka Pathak, Lauren Raymond, Rebecca Richardson, Rachael Gates, Joyce Hahn, Kelly Vosilus, and Jenni Onslow. May your doors always be open and your table always be full.

Table of Contents

1. Introduction

Water is necessary for human survival. Without adequate access to safe drinking water, disease-causing pathogens can be spread throughout populations. Many fecal pathogens cause diarrhea and are passed to children through their drinking water. Each year, there are approximately 1.8 million deaths due to diarrhea; among children under 5 years old, diarrhea is the second most common cause of death (UNICEF and World Health Organization 2009). While drinking unsafe water is one pathway of transmitting diarrhea, safe water is an important step in improving the quality of life and health of a population. It is often assumed that improved drinking water directly reduces cases of diarrhea in a community. Hence, diarrheal cases and rates are commonly used as indicators of health.

Many people make decisions daily regarding drinking water that affects the health of their households. Access to indoor piped water is a luxury for many people in rural areas of sub-Saharan Africa. Often, women rely on other sources to collect water including rivers, streams, puddles, rainwater, boreholes, wells, and community pumps. Women generally go to the source nearest their households. After carrying water back home, households decide how to store their water and whether or not to do anything to improve its quality.

1.1 Pathways of spreading disease

There are five main acknowledged pathways that spread disease-causing pathogens: drinking contaminated water, eating contaminated food, touching contaminated surfaces or hands and then ingesting harmful pathogens, and flies cross-contaminating food. A person infected with an enteric disease sheds pathogens via defecation. Through inadequate sanitation and personal hygiene after defecation, pathogens can be transferred to surfaces, food, and to drinking water. Contamination of drinking water can happen at many times in the household water cycle: at the water source, while transporting the water home, via bio-film inside the stored container, through water-dippers, and during storage.

1.2 Factors that impact the spread of disease

Pathways transfer microorganisms that can cause disease, are neutral, or have a symbiotic relationship with humans. Non-viable non-living pathogens cannot cause disease if consumed. Bacteria, protozoa, and viruses all have different survival times in the environment that determine their viability and infectious dose. An infectious dose is the number of pathogens that must be consumed in order to cause disease. Different organisms have varying infectious doses; one protozoan oocyst may cause illness in a person whereas it may take 10 bacteria to cause illness in the same person. Resistance increases the infective dose necessary to cause illness and is developed through previous exposure to the pathogen, through community immunity, and through vaccinations.

1.3 Measuring water quality

Since it is not possible to directly test for human feces in drinking water and not economically feasible to individually test for each disease-causing pathogen, indicators of probable contamination are used. Levels of *E. coli,* fecal coliforms, and thermotolerant fecal coliforms have been used to identify probable contamination of water. *E. coli* and thermotolerant fecal coliforms are the most reliable indicators and are used widely in the field. Without further serotyping of *E. coli* and thermotolerant fecal coliforms, it is inferred that these microorganisms are viable and represent risk.

Another way to measure water quality is to use what is known about the water source as a proxy for the consumed household water. The "Millennium Development Goal 7 Target C" seeks to cut the number of people in half who do not have access to safe drinking water by 2015. While the global target has been met, many households in rural sub-Saharan Africa still lack access to safe drinking water sources. The population who has access to *improved* drinking water is assumed to have *safe* drinking water. Safe drinking water is defined by the World Health Organization as meeting the microbial, chemical, and physical characteristics (World Health Organization 2011). *Improved* drinking water sources are protected springs, boreholes, protected dug wells, rainwater collection systems, public standpipes, and household water connections. Unimproved sources either do not provide water of high quality or are not affordable long-term methods of securing water.

The Joint Monitoring Program for Water Supply and Sanitation (JMP), a partnership between the World Health Organization (WHO) and the United Nation's Children's Fund (UNICEF), was established to globally advise and to store water and sanitation data. The JMP provides guidance on how to collect water quality and currently uses the *improved* or *unimproved* source type to indicate access to *safe* or *unsafe* water. The JMP's new recommendation for 2015 is to first test water at the household level for the presence or absence of *E. coli* for all future Demographic Health Surveys (DHS), Multiple Cluster Indicator Surveys (MCIS), and Living Standards Measurement Surveys (LSMS) (JMP and GLAAS 2010). Secondarily, countries are encouraged to sample at the point of delivery or collection source. This shift to quantify the probable contamination of households' drinking water was made because when compared to measuring springs' risk the probable contamination of households' drinking water best measures households' health risk (Wright, Gundry et al. 2004).

Seasonality can affect water quality. Water quality may vary depending on the time of day the water was collected and the water storage time for the area. Based on a time series study in Northern Ecuador, one-time grab samples from water sources had considerable variation (Levy, Hubbard et al. 2009). Therefore, it is reasonable to think that household water quality variation may be wide in a community. Schmidt and Cairncross suggest that smaller and medium research studies in each country will contextualize the complex issues of sampling, seasonality, clustering, and human behavior so that it will be possible to understand why losses in diarrheal disease may or may not be seen after water quality interventions(Schmidt and Cairncross 2009).

Since it is unlikely that the DHS, MCIS, and LSMS surveys will collect more than one sample from a household within one year, smaller research studies can be used to develop models for seasonality, human behavior, and clustering within a country to understand the variation seen in the current surveys. A focal research study conducted in western Kenya will be used to provide an example of how to contextualize large one-time presence absence survey samples of household water quality within a smaller longitudinal research study.

1.4 Research Gap

Many organizations, governments, and individuals are invested in finding culturally acceptable solutions for safe water at an appropriate scale to improve the health of the community. Yet, there is a gap in the existing literature: water quality is not usually measured at both the beginning and the end of a study. It is instead common to measure water quality either at the household or the source in the beginning of a study, implement an intervention, and then to conduct surveys measuring whether diarrhea rates decreased over the study period. Health scientists measure the final outcome of interest, illness due to drinking contaminated water via cases of diarrhea and malnutrition, rather than measuring the intermediate outcome of interest. These studies miss the middle step- whether or not the intervention actually improved drinking water quality. Without quantifying whether water quality is improved, it is difficult to attribute an increase or a decreased health impact to the intervention.

1.5 Intra-cluster correlation coefficient

A cluster randomized controlled trial's intra-cluster correlation coefficient (ICC) measures the relatedness of clustered data (Campbell, Grimshaw et al. 2004). The ICC, rho, is calculated by dividing the variability between clusters by the sum of the variability between clusters and the variability within clusters (Killip, Mahfoud et al. 2004; Duflo, Glennerster et al. 2007). As the ICC increases from 0 to 1, an effective sample size of a large number of clusters with fewer samples per cluster is necessary (Killip, Mahfoud et al. 2004; Duflo, Glennerster et al. 2007). The ICC of households' stored water quality explains how much of the households' water quality variability is due to the variation between village clusters versus the variability within clusters (Campbell, Grimshaw et al. 2004). Using the ICCs of households' stored water and of springs' water, power calculations and effective sample sizes will be provided for the logarithmic MPN of *E. coli* method and the presence absence test to detect differences in water quality interventions.

1.6 Objectives

This thesis has two main audiences: health researchers and professionals who will use the JMP's new guidelines to improve water quality. This project provides necessary information for development practitioners and health researchers interested in measuring water quality of clustered household users across time.

The objectives and research questions are listed below.

Objective 1: To analyze the water quality variance by cluster and by season using four water quality summary methods

1. What is the variability of households' stored water quality and springs' water quality across months of the year and across bi-weekly monitoring rounds using four water quality summary methods?
2. Of the four water quality summary methods, which is the most useful for detecting differences of water quality of household clusters and of spring clusters?
3. Is the variability of water quality between clusters greater than the variance of water quality within clusters for households and for springs?

Objective 2: To provide sample size & power calculations for future water quality studies using the focal Kenyan study as an example

1. Do the intracluster correlation coefficients for spring and households' stored water quality vary when water quality is measured on a monthly versus a bi-weekly time scale?
2. Given that the JMP recommends using presence absence tests post-2015, how does the presence absence test compare to the logarithmic MPN method of enumerating *E. coli* using the obtained intra-cluster correlation coefficients for springs' water quality and for households' stored water quality?

3. Using the ICCs of presence absence tests, what is the appropriate number of clusters and samples per cluster to collect at the village level and at the community level for future studies?

1.7 Focal study: the Spring Improvement Project Household Survey and Bi-weekly Monitoring

The focal study provides an excellent opportunity to test the JMP's new framework for measuring water quality before it is implemented in 2015. As a randomized controlled trial, the high quality data obtained from the Spring Improvement Household Survey dataset will be used in this thesis. The Spring Improvement Project Household Survey dataset (SIP-H) includes information collected annually from rural Kenyan households in the Western Province in the Busia and Butere-Mumias districts from 2004 to 2008. In a nested subset of the SIP-H, water was collected in 15 bi-weekly monitoring rounds from households and springs to test the effectiveness of providing WaterGuard at the household level. The goal of the randomized controlled trial was to test whether by improving unprotected springs gains could be seen in water quality and in health.

1.8 Kenyan context

In 2008, 52% of the rural population had access to improved water sources (12% piped and 40% other) (World Health Organization and UNICEF 2010). Over 25% of Kenyan households spend 30 minutes roundtrip or more collecting water (World Health Organization and UNICEF 2010). In the Busia and Butere-Mumias districts, 90% of

households have access to springs and 72 % of all water trips are to springs (Kremer, Leino et al. 2011). After springs, most people use sources in the following order: shallow wells, boreholes, and surface waters (Kremer, Leino et al. 2011). Drinking water is typically stored in clay pots (Kremer, Leino et al. 2011). Kenya has a short rainy season from October to November and a long rainy season from March to June which can vary slightly from year to year (Opiyo, Mukabana et al. 2007). The dry seasons are from December through February and from July through September (Opiyo, Mukabana et al. 2007).

1.9 Literature Review

1.9.1 Health Impacts attributed to poor water, sanitation, and hygiene

Poor water, sanitation, and hygiene cause a large portion of diseases especially among children under five years old. Diarrhea causes 16% of all deaths for this age group (UNICEF and World Health Organization 2009). Diarrheal diseases represent 4.1% of the total disability adjusted life years' (DALYs) burden of disease (World Health Organization 2012). Children are a highly vulnerable population to diarrhea because their immune systems are still developing. Ensuring access to and consumption of clean water is critical to address the following health issues: diarrhea, stunting, wasting, malnutrition, parasitic infections, and enteric infectious diseases (Gunther and Fink 2011). Since these conditions are often widespread in a community, essential care is not always accessed for these children. Drinking water of poor quality is one transmission route of fecal pathogens that which cause these illnesses.

1.9.2 Measuring water quality

Human Fecal Pathogens

Viruses, bacteria, and protozoa excreted from infected humans can all cause disease and have different ecological persistence times in the environment. If a person ingests a nonviable pathogen, it will not cause disease. Both humans and animals can be infected with disease and excrete pathogens in their feces. Generally, human fecal pathogens prefer human hosts and survive by spreading through the human community. People who ingest viable fecal pathogens may be asymptomatic and can easily spread disease to friends and neighbors.

1.9.3 Indicators of fecal pathogens

Using indicators of fecal pathogens

Indicators for human fecal pathogens do not directly measure contamination, but they point towards the probable presence of pathogens in the drinking water and probable contamination. If each known pathogen were tested for in a drinking water sample, this process would be expensive, time-intensive, and require a lot of water. Individual specific tests are available and useful in outbreak situations. Using a specific test can underestimate the risk of drinking the water.

Quantified indicators are used to measure probable contamination of drinking water and offer a localized picture of what a person might be drinking. These include thermo-tolerant fecal coliforms (TTC), *E. coli*, phages, and bacteroides. Phages can be used to

test for specific viruses and protozoa; these tests are rare as the methodologies are difficult, the tests are expensive, and time intensive (World Health Organization 2011).

It is acceptable to test for one of the following bacteria which should be absent from a 100 milliliter drinking water sample: *E. coli* (most reliable and preferred) or thermo-tolerant fecal coliforms (Wright, Gundry et al. 2004; World Health Organization 2011). Low (<1), intermediate (1-10), high (11-100), and very high (>100) risk levels are measured in most probable number of *E. coli* per 100 milliliters (World Health Organization 2011). If samples have levels that are below the detectable limit (<1 MPN), this does not mean the water is not contaminated with fecal pathogens. It indicates that the test was not sensitive enough to detect the low levels of pathogens (World Health Organization 2011).

Detecting E. coli

There are several ways to detect *E. coli* in water samples. A quick presence or absence test for *E. coli* can be done with 1 liter samples or smaller dilutions using the Colilert membrane filtration method; this does not provide the level of contamination of a water sample (USGS 2007; World Health Organization 2011). In the field, this test has a high sensitivity (83%), specificity (84%), and positive predictive value (92%), but also a large percentage of false negatives (70%) (Trottier 2010). Quantification is more expensive, but can be done with quantitrays that provide the most probable numbers of *E. coli* from 1 to 200 or from 1 to 2,400 using the Colilert method (USGS 2007; IDEXX 2012). With the Colilert method, water is added to a tray with small wells, separated

from each other via a sealer, incubated, and then counted to see the number of yellow wells. Yellow indicates the presence *of E. coli;* IDEXX has a table that allows a researcher to quantify the most probable number of *E. coli* present based on the number of yellow wells present. Researchers use the number of small and large yellow wells to calculate the most probable number of *E. coli* using an algorithm.

Alternatively, membrane filtration using thermotolerant *E. coli* Agar (mTEC) can be used to quantify *E. coli* colonies (Office of Water 2002). With this method, water is pulled via a vacuum through a filter; the filter is small enough to catch the *E. coli* (Office of Water 2002). The filter is transferred to a plate with medium to grow on and incubated. The number of yellow colonies, yellow-brown colonies, and yellow-green colonies are counted. Next, calculations are done based on the amount of water filtered to present the results in colonies per 100 milliliters. The mColiBlue24 method is another type of membrane filtration method where the media is already present in the dishes. Water is pulled through the filter and then the filter is placed on the blue dish in the incubator to grow overnight. It is used to quantify both *E. coli* and total coliforms. In a side-by-side comparison of 10 EPA water quality testing methods, the mColiBlue24 method did not detect high spiked levels of *E. coli* (Olstadt, Schauer et al. 2007).

Each test uses different amounts of water. Membrane filtration and IDEXX are usually performed with 100 milliliters of water. Sampled water can be diluted for both methods. In the past, researchers often used the 5 multiple tube method to detect a smaller range of *E. coli* and total coliforms. Only 1 milliliter of water can be tested for

petrifilms. After incubation, the presence of red and blue colonies indicate the presence of total coliforms and *E. coli*, respectively (Petrifilm 3M).

Detecting specific fecal pathogens

C. perfringens, a sulfide-reducing clostridia, can be detected using the Hydrogen sulfide(H_2S) method; *C. perfringens* is a bacteria present in 13-35% of human feces (Ashbolt, Grabow et al. 2001). Hydrogen sulfide producing bacteria also include *Enterobacteriaceae, Citrobacter freundii, Enterobacter, Clostridia, Escheria, salmonella, acinetobacter, aeromonas, morganella, Kliebesi, and Edwardsiella* (Trottier 2010). In a comparative field test of hydrogen sulfide, Colilert, Easygel, and Petrifilm tests, the Colilert and hydrogen sulfide tests were strongly recommended whereas the Easygel and Petrifilm gels were strongly discouraged to use as a single test for improved or unimproved water quality (Trottier 2010). Trottier found that combined tests yielded more accurate results than single field microbial tests.

1.9.4 Millennium Development Goals' safe water indicator

The current global method used by the MDGS to monitor drinking water quality is to report the proportion of people in a country who have access to *improved* drinking water. *Improved* drinking water quality is a proxy for *safe* drinking water. Improved sources are protected springs, boreholes, protected dug wells, rainwater collection systems, public standpipes, and household water connections. Unimproved sources are unprotected springs, unprotected wells, bottled water, vendor-provided water, tanker truck water, and surface waters. Improved water sources are considered to have better

overall water quality and affordable access than unimproved water sources. Counting the proportion of people who have access to sources that are improved or unimproved is a step removed from taking water samples and quantifying the probable contamination. However, measuring for improved or unimproved water sources is far less expensive and time intensive.

Determining access to safe water

Currently, 84% of the 1.1 billion people who do not have access to improved drinking water sources live in rural areas (World Health Organization and UNICEF 2006). It is important to know how this statistic was produced. To determine the proportion of people that use *improved* versus *unimproved* water sources, a sample of households is asked through DHS, MCIS, and LSMS surveys about their water collection practices. Then, the sample is weighted based on the overall rural or urban population numbers and on the amount of expected clustering of samples from each community. Weighting via clustering is necessary because people living in one area are more likely to be similar to each other than people living in areas that are farther apart. If everyone from a village uses unprotected springs and river water, then fewer samples are necessary from the village. However, if there are many types of sources accessible within the same distance, then a greater number of households need to be sampled to find out which sources are used most often and the proportion of households that use them. Next, one or two summary statistics for the entire nation are presented: national improved access to drinking water and rural versus urban improved access to drinking water.

This process of creating a summary statistic is necessary since the entire population from a country was not surveyed. Regional and worldwide statistics provide a large-scale picture of the work that is ongoing and target areas. 56% of people living in sub-Saharan Africa have access to improved drinking water sources, but this is far below the global average of 78% who have access to improved sources in rural areas (World Health Organization and UNICEF 2006; JMP and GLAAS 2010). Access to improved drinking water sources in rural sub-Saharan Africa is a priority.

1.9.5 Presenting statistics

Once the most probable numbers of *E. coli* have been determined (<1 to 2419+), there are different ways to present the results. If large numbers of *E. coli* values are below or above the detection limits, then histograms displaying the distribution of *E. coli* will be right skewed. The WHO recommends using arithmetic means of *E. coli* levels so as not to underestimate the drinking water risk or their categories of drinking water (World Health Organization 2011). Yet, microbiologists and health scientists often used log-transformed *E. coli* values and geometric means to make recommendations. The geometric mean is used for right-skewed data as an unbiased sample of log values of the median *E. coli* values (Helsel and Hirsch 1991). Log levels of *E. coli* are used to measure the reduction capacity of a water improvement method.

1.9.6 JMP's New Guidelines

The JMP issued the following recommendations after a technical meeting to develop a plan for post-2015 (JMP Technical Task Force 2010):

1. Countries should use the pass or fail measure of *E. coli* for 0 cfu/100mL as it is the best available indicator of fecal contamination; *quantifying E. coli* levels is preferred, but only if results are independently reviewed;
2. Priority should be to first take samples at the household or point of use level; if possible, also sample sources and points of delivery; and
3. Priority should be to first take samples for arsenic and fluoride; other metal contaminants may be country specific and fall within the discretion of the country.

The JMP recommends adding water quality sampling to future Demographic Health Surveys, Living Standards Measurement Study surveys, and Multiple Indicator Cluster Surveys. This will result in taking grab samples primarily during national dry seasons. National water quality averages will not accurately represent *E. coli* levels over time. The JMP recognizes that the microbiological quality of improved sources often does not meet the WHO's criteria of less than one colony of *E. coli* per 100 milliliters (JMP and GLAAS 2010).

1.9.7 Source variation

Water quality varies depending on the type of source used. In rural areas, communities rely heavily on rivers, lakes, springs, earth pans, rainwater, and boreholes. Typically,

boreholes and protected springs have higher water quality although they can still be contaminated (Nussbaumer 2008; Pickering, Davis et al. 2010). Collected rainwater quality depends greatly on the type of collection container and whether it is accessible to fecal contamination from birds, sediments from leaves and air particulates, the type of roof, and how households access the water (World Health Organization 2011). Particulates from thatched roofs can run off into the water; if the roof has a bitumen based coating, the slightly acidic rain water can dissolve high quantities of metals into the drinking water (World Health Organization 2011). Generally, rainwater quality improves throughout the rainy season (World Health Organization 2011).

1.9.8 Seasonality

While available water quantity is a factor affecting the decisions households make (Ahuja, Kremer et al. 2010), it often means households collect drinking water from high risk sources. Water quality interventions may be more effective than previously described (Fewtrell, Kaufmann et al. 2005; Clasen, Roberts et al. 2007). The majority of studies are framed to measure the reduction of diarrhea morbidity and mortality and use *E. coli* levels to measure exposure to fecal pathogens. In a recent review, water quality interventions lasting for less than 12 months (0.56, CI: 0.47-0.66) had a stronger effect than either water supply (0.82, CI: 0.71-0.96) or water quality interventions which lasted for 12 months or more (0.81, CI: 0.67-0.97) (Waddington, Snilstveit et al. 2009).

Rainfall can impact the microbiological quality of water (World Health Organization 2011). In Waddington's systematic review (Waddington, Snilstveit et al. 2009), the literature is mixed as to whether interventions have a larger effect during the rainy season (Aziz, Hoque et al. 1990; Ahmed, Zeitlin et al. 1993; Luby, Agboatwalla et al. 2006) or during the dry season (Jensen, Ensink et al. 2003; Stauber, Oritz et al. 2009; Tiwari, Schmidt et al. 2009). In Cambodia, significant water quality and variability decreases were found in the levels of *E. coli* and total coliform contamination at both open and shallow wells (Bennett, Shantz et al. 2010). The median *E. coli* contamination level was $10^{3.2}$ CFU per 100 milliliters during the rainy season (Bennett, Shantz et al. 2010).

1.9.9 Human Behavior

The specific point in time at which a sample is collected affects which conclusions can be drawn regarding the household's water quality. Water samples can be taken from households when a researcher visits, at the collection source from a household's water jug, or during the day at each time a person takes a drink of water. Collecting water from a household's jug presents a general picture at one point in time of potential contamination whereas sampling each time a person in the household takes a drink of water presents a closer representation of what the individual is drinking. Multiple samples within the same household could change throughout the day as a family uses the water.

Over time, microbes settle out in a container, die, and grow. Children may put their hands in the water; if their hands have fecal matter on them, this is spread to the water. Using a dipper or cup to scoop water out of a wide-necked container present the same potential of contamination. As the household pours their water for the researcher each time they take a drink, the household has less water than they usually do for drinking. This may result in household members gathering more water and storing it for less time, or in households drinking less water while researchers are present. This may bias the water quality in either direction: poorer water quality may be detected if the microbes are not allowed to settle out over time. Higher water quality may be detected if the water was just collected from an improved source and is not given time to be mixed thoroughly with other pathogens residing in the storage container or on children's hands.

Often, health scientists advocate for educational programs that can lead to behavior change. If households were provided with the results of their source water quality or their household drinking water quality would this change their behavior? Information regarding whether or not the source water quality was contaminated increased the use of free water treatment products in Kenya by 12-24%, but being provided with whether or not the household's stored water was contaminated did not increase the use of distributed chlorination products (Luoto 2009). Knowledge of source water quality leading to behavior change has also been demonstrated by researchers in other areas (Madajewicz, Pfaff et al. 2007; Jalan and Somanathan 2008) but it does not always translate to product uptake (Pinfold and Horan 1996; Pattanayak, Poulos et al. 2007;

Luby, Mendoza et al. 2008). Since 79% of respondents asked in Western Kenya list dirty or bad water as the most common cause of diarrhea (Dye, Apondi et al. 2011), a household may form its risk perception of source water quality cleanliness using the type of activities that they know occur at each source.

1.9.10 Relevant Studies

Water quality is rarely measured taking seasonality into account for rural areas of developing countries. This is due to a number of factors including lab constraints, remoteness of village, funding, donor will, logistics, and capacity. Fewtrel et al's meta-analysis shows that the best water quality studies conducted before 2003 only lasted for 9 months, far too short to capture seasonality effects (Fewtrell, Kaufmann et al. 2005).

Studies were included into this current review from both Wright and Gundry 2004 and from Fewtrell et al's 2005 meta-analysis based on the following criteria:

1) Studies were conducted in a rural area of a developing country

2) *E. coli, total coliforms, or fecal coliforms* were reported as an outcome **or**
   intermediate outcome measure or collected more than once from the same household or source and

3) Studies were not conducted during an outbreak response or disaster. Retrospective case control studies were excluded for two reasons: first, water quality measurements were rarely taken and second, water quality measurements of today's water do not reflect the household's water quality or risk in the past. Several selected

relevant studies conducted afterwards were also selected (see Table 1 in the appendix). Each study is summarized and described in the appendix; the conclusion is below.

Table 1: Studies included in the Literature Review

Study[1]	Water quality taken after baseline?	Water quality outcome ?	Health outcome	Reason for inclusion or exclusion
Studies from Fewtrell et al				
Alam et al 1989	None	No	Diarrhea episodes	No WQ measure
Aziz et al 1990	None	No	Diarrheal morbidity	No WQ measure
Colwell et al, 2003	% using sari filter	No	Cholera cases	No assay
Colwell et al, 2010	% using sari filter	No	Cholera cases	No assay
Conroy et al 1996	No	No	Diarrheal episodes	No WQ measure
Daniels et al 1990	No	No	Household characteristics	Retrospective study
Esrey et al 1988	No	No	Child infection of bacteria, viruses, & parasites	No direct WQ measure
Haggerty et al	No	No	Diarrhea morbidity & mortality	No WQ measure
Hoque et al 1996	No	No	Diarrhea morbidity, CFUs on mother's fingers	No WQ measure
Jensen et al 2003	*E. coli,*	Yes	Diarrhea episodes, water quality	Included
Joyce	*E. coli*	Yes	Water quality, reduction of viable *E. coli*	Included
Kirchhoff et al 1985	*E. coli* & rotavirus	Yes	Diarrheal days	Included
Iijima et al 2001	*E. coli*	Yes	Diarrheal cases	Included
Mahfouz et al 1995	*E. coli*	Yes	Diarrhea cases and stool samples	Included
Mertens et al 1990	Fecal coliforms	Yes	Diagnosis of Salmonella, shigella, *E. coli,* Camplybacter, trophozoites, rotavirus	Retrospective study

[1] All studies shown were conducted in rural areas except for Musa et al (mixed areas) and Molbak et al (rural and slum areas). Only rural studies are shown in this table although both reviews included urban studies.

Messou et al 1997	No	No	Diarrhea cases	No direct WQ measure
Nanan et al 2003	No	No	Diarrhea cases	Retrospective study
Pinfold and Horan 1996	None	No		No WQ measure
Ryder et al 1985	*E. coli*	Yes	Diarrhea cases	Location: island
Tonglet et al 1992	No	No	Factors of Buruli uclers	Retrospective study
Torun, 1982				Study in Polish
Wang et al 1989	Yes	Unclear	Diarrhea cases	Unclear WQ statistic
Wilson et al 1991	No	No	Kids' illness	No direct WQ measure
Xiao et al 1997				Study in Chinese
From Wright et al 2004				
Musa et al 1999	Fecal coliforms	Yes	Diarrhea district prevalence	Included
Molbak et al 1989	Enterobacteria in food and water	Yes	Diarrheal cases	Included
Aquapol				Cannot access
El Attar et al 1982	*E. coli* and fecal streptococci	Yes	Water quality	Included
Morin et al 1990	Fecal coliforms	Yes	Water quality	Included
Pinfold 1990	*E. coli*	Yes	Water quality, hand hygiene	Included
Simango et al 1992	*E. coli*, salmonella, camploybacter, shigella, aeromonas, yersinia enterocolitica	Yes	Water and food quality	Included
van der Hoek et al 2001	*E. coli*	Yes	Diarrhea cases	Included
Austin 1994				Cannot access
Blum et al 1990	Fecal coliforms & fecal streptococci	Yes	Dracunculiasis and diarrheal cases	Included
Heinanen et al 1988	Fecal coliforms	Yes	Water Quality	Included
Kefauver 2000				Cannot access
Knight et al 1992	Fecal coliforms	No	Diarrheal associations	Retrospective study
Lehmsuluoto 1986	Fecal coliforms	No	Water quality	Included

Platenburg and Zaki 1993	Fecal coliforms	Yes	Source water quality	Included
Rajasekaran et al 1977	Shigella	Yes	Diarrhea cases	Included
Shears et al 1995	Fecal coliforms	Yes	Microbial resistance	Included
Sutton & Mubiana 1989	Fecal coliforms	Yes	Water quality	Included
Trevett et al	Fecal coliforms	Yes	Diarrhea cases	Included
Verweij et al 1991	Fecal coliforms	Yes	Skin infections	Included
Woodhouse 1990				Not a study
Empereur-Bissonnet et al 1992	Fecal coliforms, total coliforms	Yes	Water Quality	Included
Shiffman et al 1978	Coliforms	No	Diarrhea, Food wastage	Unclear statistics
Feachem 1978	*E. coli*, fecal streptococci	Yes	Water quality, cases of disease	Included
Kaltenthaler et al 1996	Fecal coliforms	Yes	Water quality	Included
Lindskog & Lindskog 1988	TC, FC, fecal streptococci	Yes	Water quality (fecal coliforms	Included
Mazengia et al 2002	Turbidity & fecal coliforms	Yes	Water quality	Included
Chidavaenzi et al 1998	Fecal coliforms, total coliforms	Yes	Water quality	Included
Esrey et al 1986	*E. coli*, giardia, camplyobacter	Yes	Water quality, diarrheal cases, child growth	Included
Sandiford et al 1989	Fecal coliforms	Yes	Drinking water quality	Included
Tomkins et al 1978	Fecal coliforms	Unclear	Malnutrition, wasting, stunting	Included
Young & Briscoe 1988	Fecal coliforms	Yes	Diarrhea	Retrospective study
Other Relevant Studies				
Levy & Nelson 2008	*E. coli,* enterococci	Yes	Water quality	Included
Levy & Hubbard 2009	*E. coli*	Yes	Water quality	Included

1.9.11 Conclusion

A surprising number of water quality studies have a small sample size or leave out crucial information necessary to judge the internal and external validity of the study.

Only half of the 32 studies included clearly described their sampling methods and sample size (Feachem, Burns et al. 1978; Heinanen, Chandiwana et al. 1988; Lindskog and Lindskog 1988; Molbak, Hojlyng et al. 1989; Sandiford, Gorter et al. 1989; Sutton and Mubiana 1989; Esrey, Potash et al. 1991; Empereur-Bissonnet, Salzman et al. 1992; Platenburg and Zaki 1993; Mahfouz, Abdel-Moneim et al. 1995; Shears, Hussein et al. 1995; Kaltenthaler, Drasar et al. 1996; Cidavaenzi, Jere et al. 1998; van der Hoek, Konradsen et al. 2001; Mazengia, Chidavaenzi et al. 2002; Levy, Nelson et al. 2008; Levy, Hubbard et al. 2009). Of these, studies took source samples from a range of 2 to 320 sources. Across these studies, water quality was sampled from 25 to 703 households within a study.

A handful of excellent studies have been conducted to observe households' water quality and source water quality across time. Levy's prospective case control study discovered that river water quality has a high hourly variability(Levy, Hubbard et al. 2009). Feachem and colleagues concluded that transportation of water increased fecal coliforms from the source to the household by 103 coliforms per 100 milliliters(Feachem, Burns et al. 1978). Rajasekaran et. al measured source and household water quality twice a month; they found higher levels of *Shigella* in the monsoon season and levels above 10 MPN of *E. coli (Rajasekaran, Dutt et al. 1977)*. This project aims to add to the body of literature.

2. Methods

2.1 Spring Improvement Project

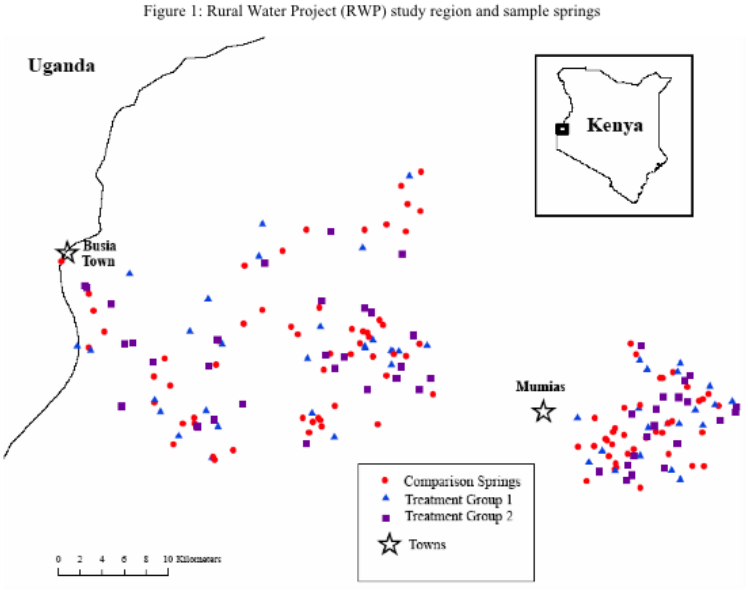
The Spring Improvement Project is a cluster randomized controlled trial designed to measure the community and household level health gains achieved by protecting spring water sources. The team led by Michael Kremer, Jessica Leino, Edward Miguel, and Alix Peterson Zwane found a 66 percent reduction in fecal contamination of protected spring water and a 23 percent reduction in fecal contamination of household water due to spring protection (Kremer, Leino et al. 2011). Ahuja and Kremer et al provide an in-depth discussion of health gains from spring improvement (Ahuja, Kremer et al. 2010; Kremer, Leino et al. 2011).

Households' stored water quality and springs' water quality was sampled throughout the study. At baseline, the team observed higher household water quality than source water quality (average difference of 0.51 log levels of *E. coli)* and speculated that this may have been due to each household using multiple water sources (Kremer, Leino et al. 2011). In the third survey round, the team observed that unprotected springs (3.6) had higher water quality than boreholes (4.1), shallow wells (5.2), lakes and ponds (6.0), and rivers and streams (7.0) measured by log levels of MPN of *E. coli* (Kremer, Leino et al. 2011). This finding is different from other water quality studies that show boreholes have the highest water quality. Two main water sources besides a protected spring in each community were sampled to obtain this information. It is possible that throughout the project villagers used protected spring water more often than

unprotected spring water; therefore this would reduce the number of people and potential contamination at the unprotected springs.

Busia is located near the Ugandan Kenyan border (Figure 1). Butere-Mumias is southeast of Busia and north of Lake Victoria.

Figure 1: Spring Improvement Project study region



(Kremer, Leino et al. 2007)

2.2 Spring Selection

Springs were eligible to be selected into the study on four conditions: if the spring was unprotected, within the Busia or Butere-Mumias districts of Kenya, not seasonally dry, and if no known contamination was present upstream (Kremer, Leino et al. 2011). From this selection criteria, 200 springs were randomly selected into the study; after baseline surveys, researchers determined that 16 randomly assigned springs were not eligible to be included in the study (Kremer, Leino et al. 2011). Therefore, 184 springs

were used for the study. For further discussion of sample selection see Kremer *et al.*'s supplementary online appendix (Kremer, Leino et al. 2011).

2.3 Household Selection for SIP-H Study

At each spring, enumerators asked individuals to provide the names of all households that use the spring. Next, enumerators visited 3 to 4 households near the spring to obtain a list of all households that use the spring. Households were placed on an eligible selection list into the spring improvement project if their name appeared on both lists; 7 to 8 households from the eligible list at each spring were randomly selected as representative households (Kremer, Leino et al. 2011).

2.4 SIP-H Survey Rounds

The spring improvement project consisted of a baseline survey, three monitoring surveys, and 15 rounds of bi-weekly monitoring surveys. From August 2004 to February 2005, a baseline survey was conducted and water quality samples were taken from 184 springs and 1,384 households (see Table 2) (Kremer, Leino et al. 2011). The first survey round was conducted from April until August in 2005 and water quality samples were collected from 175 springs and 1,250 households (Kremer, Leino et al. 2011). The second survey round was conducted from August until November of 2006; water quality was tested from 183 springs and 1,283 households (Kremer, Leino et al. 2011). The third survey round was conducted from January through March of 2007 and consisted of 184 springs and 1,231 households(Kremer, Leino et al. 2011). After the

third round of improvement, 15 rounds of bi-weekly monitoring water quality samples were taken from a subsample of 321 households (Kremer, Leino et al. 2011).

Table 2: SIP Spring & Household Selection					
Round	Sampling Timeframe	Springs sampled	Unprotected springs[2]	Households Sampled	Households using unprotected springs[3]
Baseline	August 2004 – February 2005	184	193	1,384	1,455
Round 1	April – August 2005	175	137	1,250	944
Round 2	August – November 2006	183	91	1,283	677
Round 3	January – March 2007	184	91	1,231	652
BWM rounds	May 2007 – May 2008	73	72	321	251

2.5 SIP-H Interventions

2.5.1 Spring protection

After conducting the baseline household survey from August 2004 through February 2005, springs were randomly assigned to either be improved or to remain as unprotected springs for the duration of the study. Springs were protected in three waves from January to April of 2005, August to November of 2005, and in July 2007 (see Figure 2). Following the completion of the study all remaining unprotected springs were protected (Kremer, Leino et al. 2011).

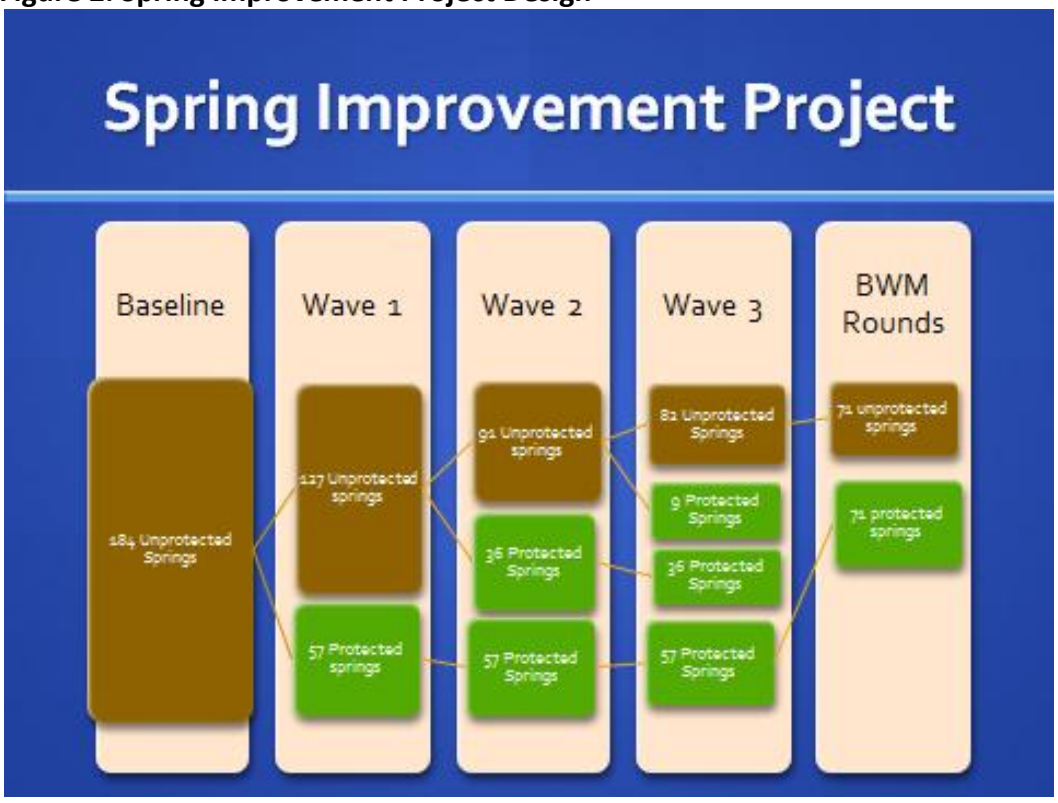
[2] The number of unprotected springs is slightly higher than the total number of springs sampled. This is because some unprotected springs were excluded after the baseline sampling when it was discovered that conditions did not meet the initial inclusion criteria.

[3] Similar to the number of unprotected springs, the number of households using unprotected springs at baseline is slightly higher than the total number of households sampled. These household samples were excluded after the team noticed that conditions did not meet the initial study inclusion criteria.

2.5.2 WaterGuard

After the second wave of spring protection, 300 of the remaining 600 households using unprotected spring water were randomized and given WaterGuard to chlorinate drinking water at the point of use.

Figure 2: Spring Improvement Project Design



2.6 Criteria for samples used in this analysis

To look at the background variation in stored untreated water, three selection criteria were used for all subsequent analyses.

1. All households that used unprotected springs were included in the analysis.
2. All households who self-reported chlorinating their drinking water were excluded from the analysis.

3. Samples that were held longer than 6 hours before being incubated were excluded due to unknown differing rates of growth and die-off.

By including only non-chlorinating households that use unprotected springs, it will be possible to examine the water quality variation due to human behavior and seasonality. As the intervention gradually protected all springs, the number of households using unprotected springs decreased over time. Since households were asked about chlorination in the first monitoring round, all observations from households using unprotected springs in the baseline survey were included. Chlorination is low as self-reported chlorination in the first monitoring round consisted of 6 percent of all households.

2.6.1 Bi-weekly Monitoring Sample (BWM sample)

The bi-weekly monitoring samples will be used for most of the analysis in the project as repeated measures of water quality from the same households and springs were taken over time. 228 households were included in the bi-weekly monitoring sample for this study using the above criterion of households that used unprotected springs, were not given WaterGuard, and did not self-report using chlorine. The bi-weekly monitoring surveys consisted of a range of 33 to 141 households per round with an average sample of 69 households. These households used 71 unprotected springs (see Table 2).

2.6.2 Sub-sample

The sub-sample consists of 1,173 households and 188 springs at baseline, 641 households and 127 springs at round one, 524 households and 91 springs at round two, and 235 households and 82 springs at round three (see Table 3).

Table 3: Viable samples from unprotected springs & non-chlorinating unprotected spring household users

	Sampling Timeframe	Included unprotected springs	Households using unprotected springs
Baseline	August 2004 – May 2005	184	1,173
Round 1	April – August 2005	127	641
Round 2	August – November 2006	91	524
Round 3	January – March 2007	82	235
BWM rounds	May 2007 – May 2008	71	228

2.7 Data

2.7.1 Survey data

At baseline, enumerators collected all the different sources which a household used in each season, reported boiling and chlorination practices, methods of cleaning water storage containers, and household socioeconomic status. Respondents named, classified, and ranked all water sources they used in the dry season and in the rainy season. If households reported chlorine use, then enumerators tested samples for the presence and quantity of chlorine that was also verified in the lab.

During the baseline and first monitoring surveys, it was protocol to ask to see the household's stored water container. While standing next to the water container,

enumerators asked to sample the water. If a dipper was present, this was used to transfer water into a sterile 250 ml bottle. The survey protocol allowed the enumerator to either collect this water on his or her own or to ask for a volunteer to collect the water. There are two assumptions with this collection process: the stored water came from the main spring that households reported using and both enumerators and households used the dipper in the same way. Enumerators may have submerged the dipper lower in the clay pot than the interviewee would have drawing more contaminated water which was closer to the bottom. Enumerators may also have skimmed water off of the top possibly reducing the most probable number of *E. coli* counts.

Survey protocol in the second and third monitoring surveys allowed the interviewee to take the 100 ml storage bottle, go to the household's stored water containers, and bring a sample back. Since the enumerator did not see the water collection process, the household member may have touched the inside of the bottle with a dipper or with his or her hands.

In the third monitoring round, if children in the household primarily used one drinking water container, then the water sample was collected from that container. Otherwise, the household's main drinking water container was sampled. Households specified which water source the stored water was collected from and enumerators recorded the time of collection. Questions about source of water in the third round only referred to

sources visited within the past week. While this collection process may have been biased, it closely approximated everyday household behavior.

Due to the change in how enumerators asked for household water samples in the second and third monitoring rounds, it is important to analyze water quality data separately across rounds. This is necessary because it is not possible to know whether the observed differences are due to seasonal changes or to the way in which the questions were asked.[4]

2.7.2 Lab data

In each round, enumerators visited the households in a community to complete survey information and to collect water samples; afterwards, they collected samples from the springs. Unprotected spring samples were collected in aseptic 250 ml bottles dragged through the center of the spring; protected spring samples were taken from the water outflow (Kremer, Leino et al. 2011). Ice packs were used to store samples until processing. Samples were processed at Busia District Hospital on the night shift with IDEXX quantitrays 2000 using the Colilert method adapted from the EPA's Colilert Quantitray 2000 Standard Operating Procedures(IDEXX 2012) (Kremer, Leino et al. 2011). Most probable numbers of *E. coli* were recorded with values ranging from 1 (lowest detection limit recorded as 1) to 2419 (too numerous to count) (Kremer, Leino et al. 2011).

[4] This would be an interesting aim for a future study considering that this is not a standard procedure between studies. Quantifying whether the way in which water is asked for from households yields different results would require a careful study design that collects at least two water samples from the same households in a visit using two different ways of asking across seasons.

2.8 Analytical Methods: Analysis Plan

2.8.1Summary Statistics

It is difficult to decide how to best summarize water quality. A ratio per season, per sampling date, or per year condenses a wealth of information into one number. Wright's method of analysis was particularly useful for calculating a ratio of probable contamination of household water samples compared to probable contamination of sources' water samples (Wright, Gundry et al. 2004). While arithmetic means are recommended by the JMP, health scientists more frequently use the geometric means of water quality.

It is important to be able to compare the results to past studies using the same summary method. Therefore, water quality will be examined using four analysis methods to determine the best summary method: the logarithmic MPN of *E. coli* method, the WHO categorical risk level method, the geometric means, and the presence absence test for *E. coli*. The smallest common time unit measured across the entire RWP study is month of year. Therefore, for the bi-weekly monitoring rounds, information will be presented both by month and by BWM round (see Table 4).

Table 4: Analysis Plan

	BWM Round					
Water Quality Comparison Method	Households	Month Statistical Test	Round Statistical test	Springs	Month Statistical Test	Round Statistical Test

Log MPN of E. coli	Box plot with WHO levels	Clustered robust standard errors method (F value, p value)	Clustered robust standard errors method (F value, p value)	Box plot with WHO levels	Clustered robust standard errors method (F value, p value)	Clustered robust standard errors method (F value, p value)
WHO risk levels	Bar chart	Clustered Ordinal logistic regression (Wald x^2, p value)	Clustered Ordinal logistic regression (Wald x^2, p value)	Bar chart	Clustered Ordinal logistic regression (Wald x^2, p value)	Ordinal logistic regression (Wald x^2, p value)
Geometric means of E. coli	Box plot	Clustered robust standard errors method (F value, p value)	Clustered robust standard errors method (F value, p value)	Not possible	Not possible	Not possible
Presence absence test	Bar chart	Logistic Regression (Wald x^2, p value)	Logistic Regression (Wald x^2, p value)	Bar chart	Logistic Regression (Wald x^2, p value)	Logistic Regression (Wald x^2, p value)
Sub-sample						
Water Quality Comparison Method	**Households**	**Month Statistical Test for baseline, R1, R2, R3 Round Statistical test**		**Springs**	**Month Statistical Test**	**Round Statistical Test**
Log MPN of E. coli	Box plot with WHO levels	Clustered robust standard errors method (F value, p value)		Box plot with WHO levels	Linear Regression (F value, p value)	Not applicable
WHO risk levels	Bar chart	Clustered Ordinal logistic regression (Wald x^2, p value)		Bar chart	Ordinal logistic regression (Wald x^2, p value)	Not applicable
Geometric means of E. coli	Box plot	Clustered robust standard errors method (F value, p value)		Not possible	Not possible	Not possible
Presence absence test	Bar chart	Logistic Regression (Wald x^2, p value)		Bar chart	Logistic Regression (Wald x^2, p value)	Not applicable

Separate analyses for BWM sample and sub-sample

Water quality analyses will be conducted separately for the BWM sample and the sub-sample due to the differing unit of time which samples were collected. Within the sub-sample, rounds will be analyzed separately because the sampling methodology for collecting water quality changed across rounds in the subsample and may otherwise confound the results.

2.8.2 Logarithmic MPN of E. coli

Box plots of the logarithmic most probable number of *E. coli* of households' stored water quality and of springs' water quality will be presented by month and by round for the BWM sample. For the sub-sample, the logarithmic MPN of *E. coli* box plots will be presented by month and by round. The season of the year will be represented by the different colored months: first dry season (purple: December through February), long rainy season (blue: March through July), second dry season (black: August), and the short rainy season (orange: September through November). On each graph, the WHO risk categories will be presented by red lines as a visual comparison of the levels.

In order to test the significance of the month for households' stored water quality using the logarithmic MPN of *E. coli* method in the BWM sample and the sub-sample, the clustered robust standard errors method will be used. The regress command will be used with the cluster option to group each household's water quality observations as the households' observations are not independent of the spring (Statistical Consulting Group).

regress hhlogmpn i.month if((chlorine!=1)&(bwm_round<16)), cluster(bwmspringid)

The command "i.month" in Stata separates each month into a binary variable, creates a dummy variable, and drops one month out to use it to test against the other months in the model (Statistical Consulting Group). The same clustered robust standard errors method will be used to test the significance of the round for households and springs in the BWM round and for households in the sub-sample. Since springs' water quality was

sampled once per year and since the sub-sample is analyzed separately, a non-clustered linear regression function will be performed on spring water quality for each round.

regress springlogmpn i.month if(surveyround==0))

The F statistic and p-values will be reported for all tests of logarithmic MPN of *E. coli*. In the example using households, the F statistic measures the change in households' water quality across the months of a survey round. If the p-value is less than 0.05, then the month of the year explains some of the variability in springs' water quality.

One advantage of the clustered robust standard errors method is that the standard errors are slightly larger than those obtained using multi-level modeling, survey method modeling, or normal linear regression (Statistical Consulting Group). Since it is more difficult to reject the null hypothesis, the clustered robust standard errors method is less likely than other methods to falsely reject the null hypothesis and commit a type one error.

2.8.3 WHO drinking water risk levels

The WHO risk levels for safe water will be calculated and presented via a bar chart for the BWM sample and the sub-sample by month of the year. Households' stored water quality and springs' water quality will both be presented.

To test the hierarchal significance of the WHO risk categories, ordinal logistic regression will be conducted across the months of the year and across the BWM rounds in the BWM sample. Ordinal logistic regression tests whether water quality changes

significantly across time from the low, intermediate, high, and very-high risk levels; it does not test whether there are a significant number of samples in a specific category. Ordinal logistic regression measures the change it takes to move from one category to the next recognizing that an explicit order exists to the categories. Water within the high-risk level has higher levels of probable contamination than water within the low risk level. Stata's ologit command with the cluster option will be used for households' stored water quality.

ologit hhrisklevel i. month if ((chlorine!=1)& (bwm_round<16)), cluster(bwmspringid)

The non-cluster option will be used for the springs in the sub-sample since only one measurement was taken per spring per round in the sub-sample.

ologit springrisk i. bwm_round if ((bwm_round<16))

The cluster option will be used for the springs' water quality in the BWM sample for the month year but not for the BWM round. The Wald chi-square test statistic and the associated p-value will be reported for each test.

2.8.4 Geometric means of E. coli

One water quality sample was collected per spring; a summary statistic of the geometric mean of water quality at a spring would not be meaningful. Therefore, only household water quality will be examined. The geometric means method summarizes the mean water quality of a village's households' stored water quality of households that use the same spring.

Box plots for the village's geometric means of MPN of *E. coli* of households' stored water quality will be displayed for the BWM sample and for the sub-sample. The clustered robust standard errors method will be used to test the significance of the month of the year for the BWM sample and for the sub-sample. Stata's regress command will be used with the cluster option to group each household's water quality observations. An example of the code used for the BWM sample across the month of the year for the geometric means method is below. The F-statistic and associated p-values will be reported for all geometric means MPN of *E. coli* tests. In this case, the F-statistic is a ratio of the sample variances of the geometric means MPN of *E. coli* of households' stored water divided by the number of spring clusters.

regress geomethouse i. month if ((chlorine!=1)&(bwm_round<16)), cluster (bwmspringid)

2.8.5 Presence absence tests

The presence absence method measures the percentages of households and of springs per month or per round in each sample set that has detectable levels of *E. coli* greater than 1. To conduct these tests, the MPN of *E. coli* for households and for springs was collapsed into binary presence absence tests. Collapsing the MPN method's enumeration of *E. coli* into a summary statistic yields different information than collecting hydrogen sulfide tests as the hydrogen sulfide test's parameters are slightly different than those of *E. coli* (Trottier 2010). However, the hydrogen sulfide test was not conducted in this study. Bar charts will be used to display the presence absence tests for households and springs by month across the BWM sample and across the rounds of the sub-sample.

Logistic regression will be conducted by month and by BWM round for the BWM sample to test for significance. Households' stored water presence absence tests will be clustered based on the spring id. Stata's logistic command will be used with the cluster option for households and for springs when testing the presence absence test's significance of the month of the year for the BWM round.

logistic springpa i. month if (bwm_round<16), cluster(bwmspringid)

Springs will not be clustered for the BWM round when testing the significance of the presence absence test across BWM rounds because there is only one observation per spring. For the sub-sample, logistic regression will only be conducted by month of the year. Stata's non-cluster option will be used for the presence absence test on springs in the sub-sample across months of the year for each survey round.

logistic springpa i. month if (surveyround==1), cluster(springid)

The Stata command above displays the logistic regression equation for the presence absence test for *E. coli* of springs' water quality across the months for the first survey round clustered by the spring. The Wald chi-square test statistic and the associated p-value will be reported for each test.

2.8.5 Intra-cluster correlation coefficients

Intra-cluster correlation coefficients (ICCs) are global measurements for a specific variable in a study that account for spatial clustering among neighbors (Campbell, Grimshaw et al. 2004). Two types of ICCs will be presented: ICCs for households and for springs. The ICC of households' stored water quality describes how similar one

households' neighbors are to users within the same village cluster as the first household when compared with households that are farther away. Similarly, the ICC of springs' water quality describes whether spring B's water quality is more similar to nearby springs C, D, and E in the surrounding villages than to distant springs in villages farther away. The variance between clusters, within clusters, and precision of the ICC will also be presented across rounds using the BWM sample and the sub-sample as these measures are recommended for reporting ICCs in randomized controlled trials (Smeeth and Siu-Woon Ng 2002; Campbell, Grimshaw et al. 2004).

The ICCs of the logarithmic MPN of *E. coli* method and of the presence absence test will be calculated along with the variance between clusters and within clusters to provide guidance for power calculations. The ICC of village's geometric means MPN of *E. coli* of households' stored water quality will not be calculated because this method already includes a clustered summary statistic. If an ICC were calculated using the geometric means method, it would double count the clustering effect. Similarly, the WHO risk-level method already summarizes the logarithmic MPN of *E. coli* into multiple categories; therefore, the ICC will not be calculated for this method either. Stata's loneway command will be used to calculate the ICC. In the first example, the ICC is calculated for the logarithmic MPN of *E. coli* across the BWM samples using the cluster of spring id.

xi: i.bwmspringid

loneway HHLogMPNecoli bwmspringid if ((bwm_round<16)&(chlorine!=1))

The xi: command turns the BWM spring id into a categorical variable for all springs and drops one spring in the loneway command. In the sub-sample, the spring id is used.

xi: i.spring_id

loneway springMPNecoli spring_id if(surveyround<4)

The second command turns spring id into a categorical variable for all springs and provides an ICC using samples from springs across the sub-sample survey rounds. An ICC for springs is not provided calculated for each round because only one water quality sample was taken per round.

2.8.6 Power and sample size

Graphs with appropriate power and sample sizes will be constructed for the logarithmic most probable number of *E. coli* method and the presence absence test using the calculated ICCs. If the ICC is large for households' stored water quality, this indicates that taking an additional household's stored water quality sample in an existing cluster will not be as useful as taking a household's stored water quality sample from a new cluster (Duflo, Glennerster et al. 2007). If water quality varies widely between the clusters of households, then future projects will need an increasingly larger number of clusters in order to detect differences of water quality at the household level. Wide variability within clusters can also indicate that the sampling unit per cluster was too large (Duflo, Glennerster et al. 2007).

Secondary analytical analysis was conducted using STATA 11.2. Since de-identified data was used for the secondary analysis of this project, approval from the Institutional Review Board was not necessary.

2.9 Limitations

Detection Limits

A large portion of the data was beneath the lower limit of detection, <1 MPN of *E. coli*, using the IDEXX sampling method while a small portion exceeded the upper limit of detection, >2419.6 MPN of *E. coli*. Therefore, the general distribution of water quality data was non-normal. Although there are methods to statistically estimate the water quality below the lower and upper detection limits, these were not performed.

Clustered Nature of the Data

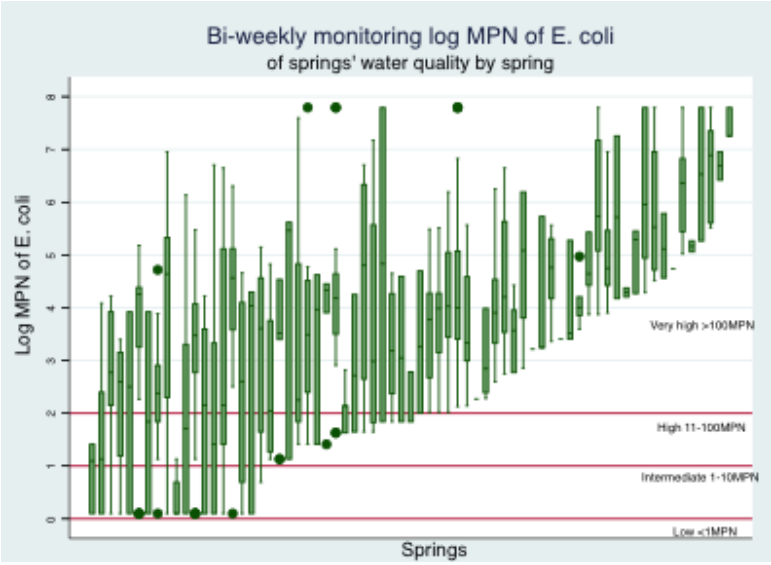
Households that are in the same village are more likely to have similar water quality than households in different villages. Households in the same village are more likely to share the same water source and to have similar household water management practices as their neighbors. Analyses will consider the clustered nature of the information, statistical limitations in summarizing data, and the round in which the data was collected.

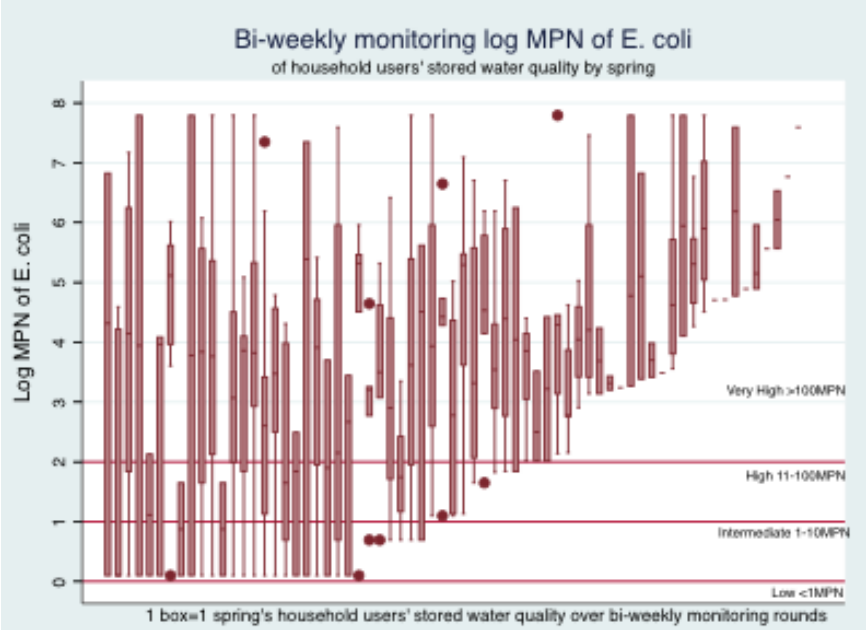
3. Results

3.1 Bi-weekly Monitoring Round Results Overview

In the bi-weekly monitoring rounds of unprotected springs, there is a wide range of water quality variability across the springs and at each spring. In Figure 3, the variability of 69 unprotected springs across 15 bi-weekly monitoring rounds is shown. Each box represents the variability at one spring across all 15 bi-weekly monitoring rounds. Some springs have narrower and higher log levels of MPN of *E. coli* whereas others have a lower and wider log levels of MPN of *E. coli*. 18 of the 69 springs (26%) had at least one observation near the WHO's low risk level. Of these springs, only 2 springs (3%) did not have observations that were above 10^2 log levels of *E. coli.* In Figure 3, households' stored water from 25 of the 69 springs (36%) had at least one observation where *E. coli* was not detected. Only two springs had all sampled household users with household stored water quality that was less than 10^2 log levels of *E. coli*.

Figure 3: BWM log MPN of *E. coli* of springs and households





A summary of the statistical tests conducted on the BWM round data is in Table 5. At the 95 percent confidence level, no statistically significant difference in household water quality was detected either by month of the year or by BWM round using any of the four methods. If standards were relaxed to the 90 percent confidence level, the presence absence test detects a statistically significant difference for household water quality by month (Wald x^2=9.17, p=0.10) and by BWM round (Wald x^2=13.42, p=0.06).

The four water quality measurement methods did detect differences for spring water quality. The logarithmic MPN of *E. coli* method detected a significant difference at the 95 percent confidence level of springs' water quality by BWM round (F=3.83, p=0.0001). At the 95 percent confidence level, the ordered WHO risk levels for drinking water were not significant for springs by BWM round or by month of the year (see Table 5). At the 95 percent confidence level, the presence absence test detected

significant differences of springs' water quality by month (Wald x^2=16.03, p=0.01) and by BWM round (Wald x^2=27.88, p=0.03).

Table 5: BWM Round[5]

	Households		Springs	
Water Quality Comparison Method	**Month Statistical Test** (F or Wald x^2 statistic, p value)	**Households Round Statistical test** (F or Wald x^2 statistic, p value)	**Month Statistical Test** (F or Wald x^2 statistic, p value)	**Springs Round Statistical Test** (F or Wald x^2 statistic, p value)
Log MPN of E. coli	F=1.11, p=0.36	F=0.82, p=0.61	F=1.73, p=0.13	**F=3.83, p=0.0001**
WHO risk levels	Wald x^2= 6.33, p=0.39	Wald x^2= 8.85, p=0.55	Wald x^2= 8.61, p=0.20	Wald x^2= 15.13, p=0.30
Geometric means of E. coli	F=0.67, p=0.67	F=missing, p= missing	Not possible	Not possible
Presence absence test	Wald x^2= 9.17, p=0.10	Wald x^2= 13.43, p=0.06	**Wald x^2= 16.03, p=0.01**	**Wald x^2= 27.88, P=0.03**

The sample size of households and of springs for the BWM sample is found in Table Y. No BWM household or spring sample were excluded from the BWM analyses. For households and springs, more than 10 samples were taken each month and more than 10 samples were taken each round.

Table 6: Sample size for the BWM round by month of the year

BWM round	Month: Households sampled	Month: Springs sampled
1	May: 111	May: 59
2	May: 105	May: 59
3	June: 69	June: 47
4	June: 90	June: 53
5	July: 52	July: 29
6	July: 56	July: 31
7	July: 13, August: 34	July: 10, August: 24
8	August: 52	August: 28
9	August: 27 September: 22	August: 14 September: 18

[5] Statistical tests conducted correspond to tests in Table 5.

10	September: 62	September: 33
11	September: 27 October: 16	September: 15 October: 11
12	October: 50	October: 30
13	October: 25 November: 8	October: 15 November: 5
14	November 45	November: 29

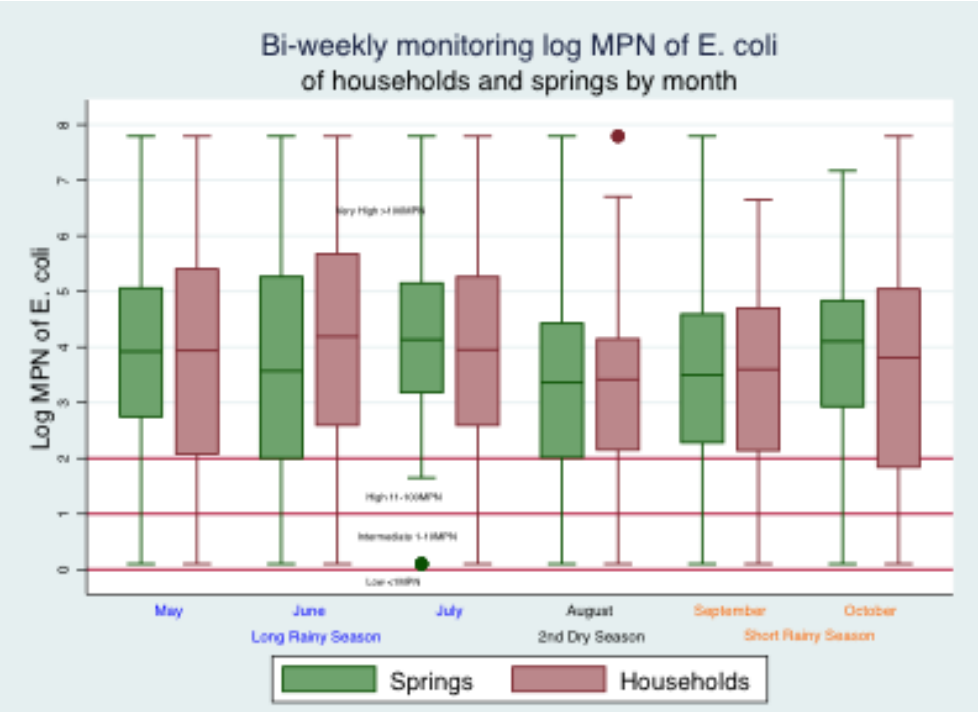
3.1.1 Seasonality measured by Monthly Variability

Logarithmic MPN of E. coli of BWM households and springs

Households' stored water quality and springs' water quality was examined using the logarithmic method of most probable number of *E. coli* (see Figure 4). It was hypothesized that the households' stored logarithmic water quality would be different across months of the year. At the 95% confidence level, a difference between the households' stored water quality was not detected across the months ($F=1.11$, $p=0.36$). The inter-quartile range for households' stored water quality across the months of the year is between log $10^{1.5}$ and log $10^{6.5}$ levels of *E. coli*.

It was hypothesized that the springs' logarithmic water quality would be different across months of the year. At the 95% confidence level, the springs' water quality was not statistically different by month ($F=1.86$, $p=0.10$). Across the months, the inter-quartile range for spring water quality stays between log 10^{2} to log 10^{6} levels of *E. coli.* In Figure 4, the inter-quartile range of households' stored water quality is tighter than the range of springs' water quality in August. Except in June, the logarithmic mean of households' stored water quality and springs' water quality is similar.

Figure 4: BWM log MPN of *E. coli*



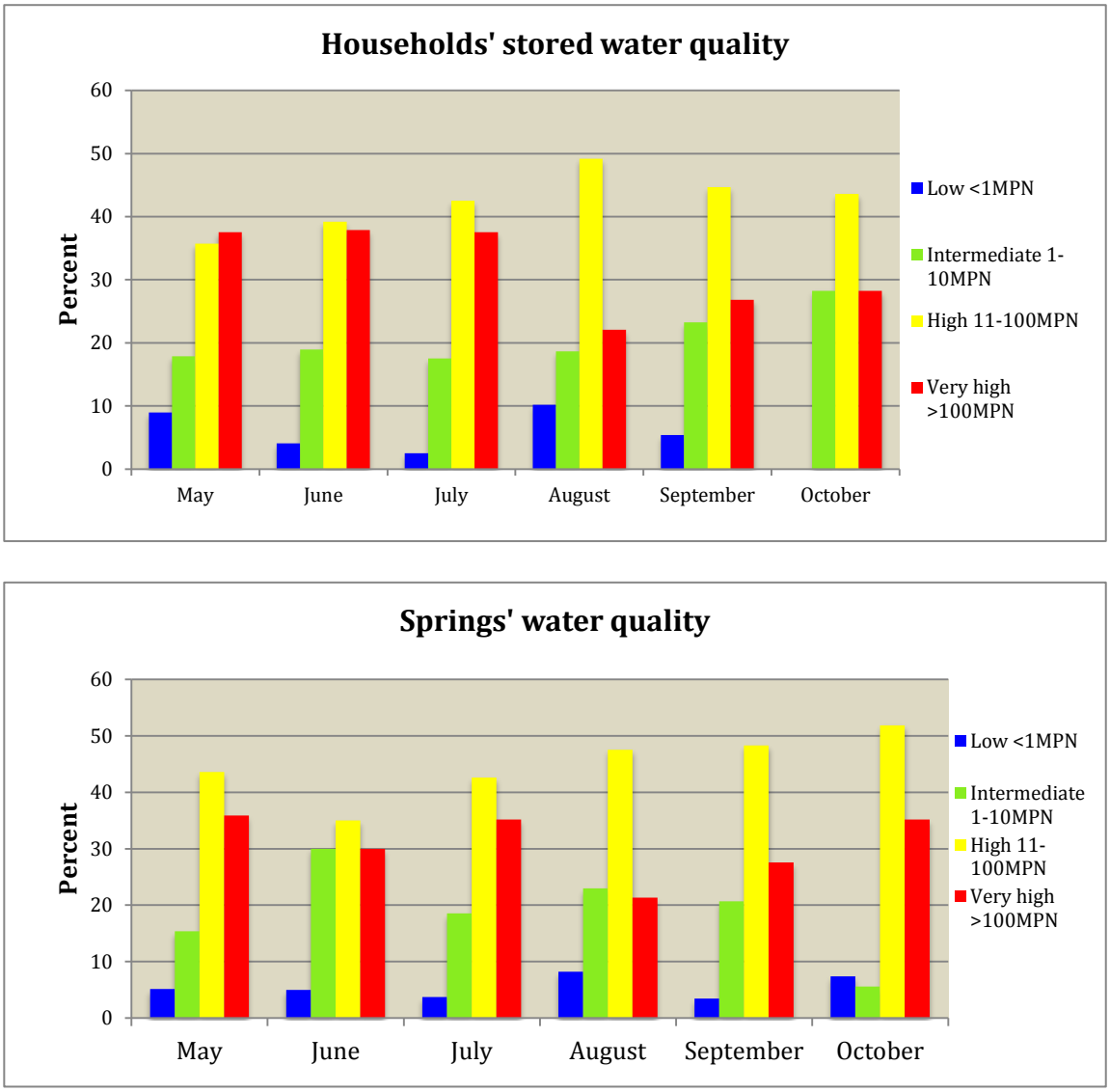
WHO drinking water risk categories of BWM households and springs

It was hypothesized that the ordered WHO's risk levels of stored drinking water would be significant across months for households' stored drinking water. At the 95% confidence level, there was no significant difference across months of household water quality detected using the ordered WHO risk categories of safe drinking water (see Table 5). In Figure 5, the percentages of households with stored water quality in the combined high and very high-risk categories varied from 69% of the households in August to 80% of the households in July. In May and August the highest percentage of households (9%) were found with stored water quality in the low risk category.

It was hypothesized that the spring water quality would differ across months of the year in the bi-weekly monitoring rounds. At the 95% confidence level, springs' water

quality was not statistically different across the months of the year (see Table 5). In Figure 5, springs' water quality in the combined high and very high risk categories varies widely by month from 66% of the samples in June to 90% of the samples in October. August (7.5%) had the highest percentage of springs' water quality samples in the low risk category.

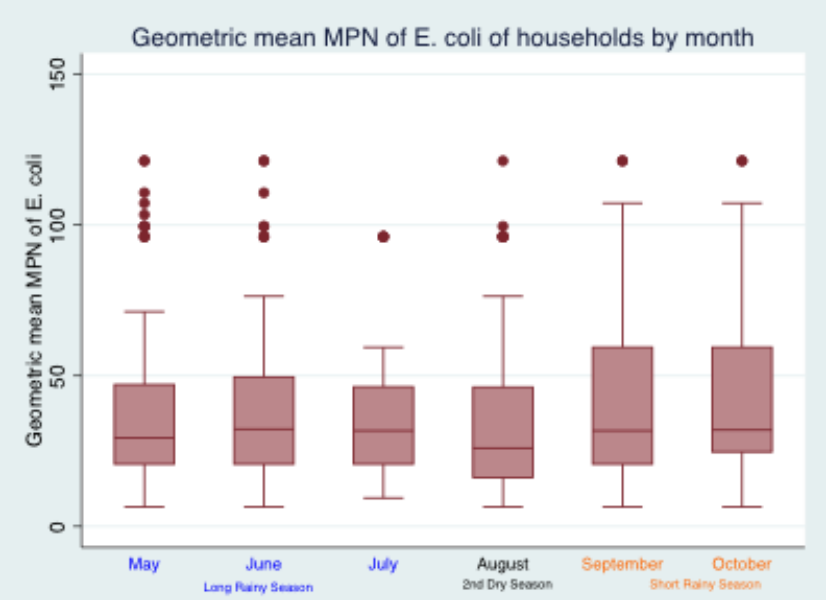
Figure 5: WHO risk level by month



Geometric means of E. coli of BWM households

It was hypothesized that the villages' geometric mean of households' stored water quality would be different across the months. At the 95% significance level, the villages' geometric mean of households' stored water quality was not statistically different across the months of the year (F=0.67, p=0.67). In Figure 6, the inter-quartile ranges for September and October of geometric means is slightly wider than that of other months.

Figure 6: Geometric mean MPN of *E. coli* by month



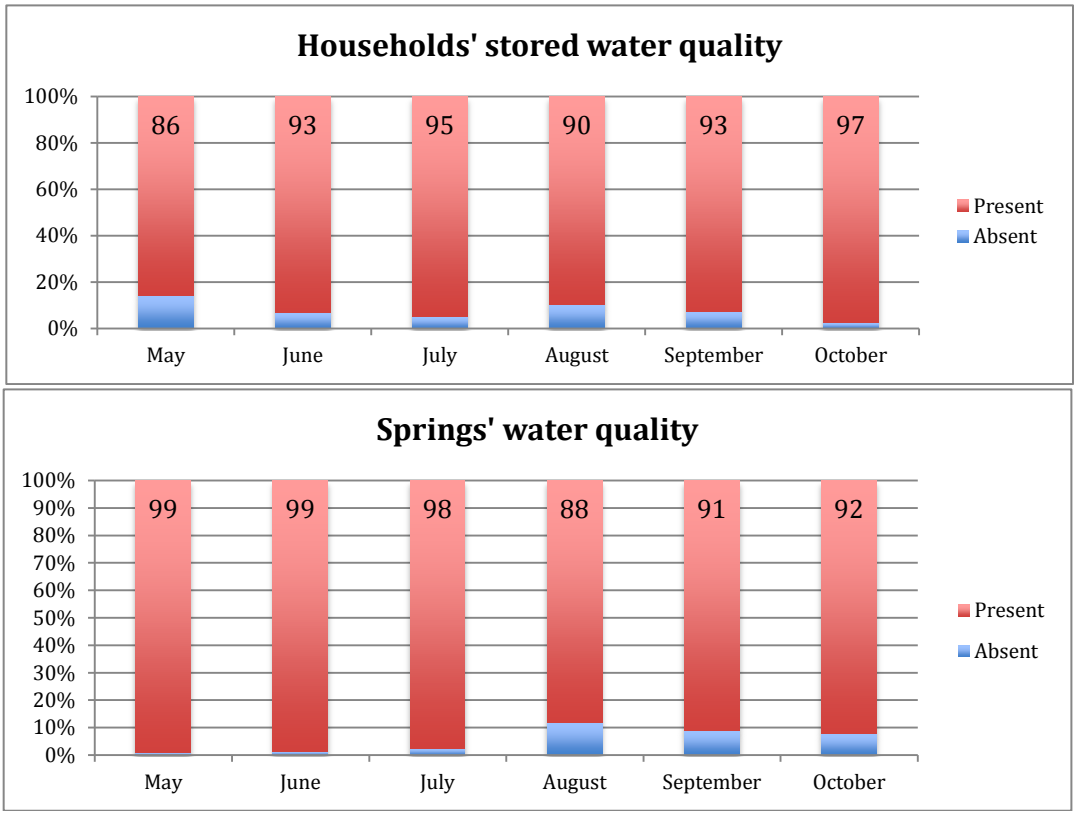
Presence absence test for E. coli of BWM households and springs

The presence absence test of monthly water quality was conducted for households' stored water and for springs' water samples (Figure 7). It was hypothesized that the presence absence test for households' stored water would not be different across the months of the year. At the 95% confidence level, the presence absence *E. coli* test of households' stored water quality was not statistically different across the months of the year (Wald x^2=9.17, p=0.10). Across the months, *E. coli* was absent from nine percent

(n=34) of all the households' stored water quality samples (n=390) with the best household stored water quality in May (14%, 16 out of 112 households) and in August (10%, 6 out of 59 households).

It was hypothesized that the presence absence test for springs' water samples would not be different across the months of the year. At the 95% confidence level, the presence absence test of springs' water was statistically different across the months of the year (Wald x^2=16.03, p=0.01). Across the months, *E. coli* was absent from five percent (n=18 samples) of the springs' 362 samples. The best spring water quality was in August (12%, n=7 of 59 springs), September (9%, n=4 of 56 springs), and October (3%, n=3 of 39 springs) with no detectable levels of *E. coli*.

Figure 7: Presence absence *E. coli* test by month



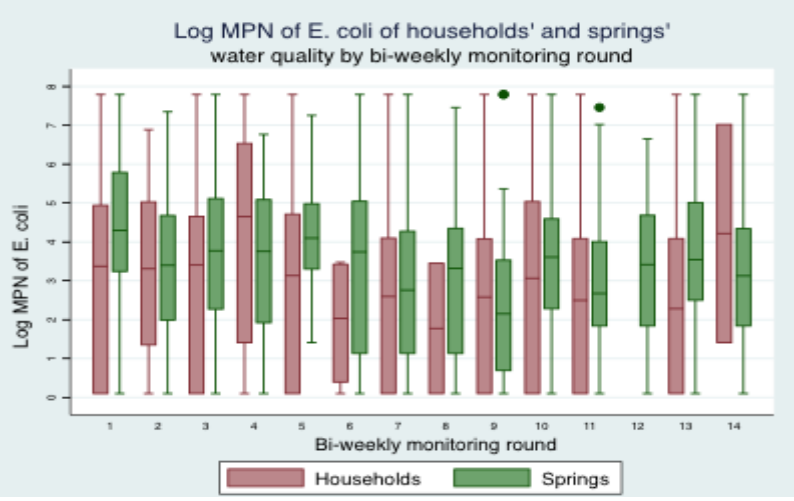
3.1.2 Seasonality measured by Bi-weekly Monitoring Round Variability

Logarithmic MPN of E. coli by BWM round of BWM households and springs

It was hypothesized that the logarithmic most probable number of *E. coli* of households' stored water would change over the BWM rounds. At the 95% confidence level, a statistically significant difference was not detected in households' stored water quality over BWM rounds (F=0.82, p=0.61). In Figure 8, the variability of households' stored water is wide as measured by the inter-quartile range of log 10^0 to log 10^7 levels of *E. coli*.

It was hypothesized that the logarithmic MPN of *E. coli* of springs' water would change over the BWM rounds. At the 95% confidence level, the springs' water quality was statistically different across BWM rounds (F=3.83, p=0.0001). The variability of springs' water quality ($10^{0.5}$ to 10^6 log levels of *E. coli*) is narrower than the variability of households' stored water quality (10^0 to 10^7 log levels of *E. coli)*.

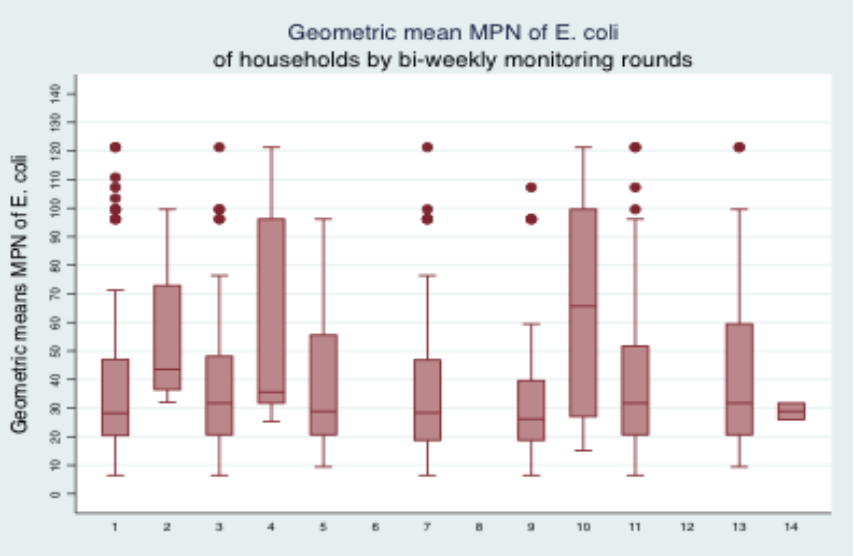
Figure 8: Log MPN of *E. coli* by BWM round



Geometric means of MPN of E. coli by BWM round of BWM households

It was hypothesized that the villages' geometric means of MPN of *E. coli* of households' stored water quality would be different across BWM rounds. STATA did not return an F test statistic or p-value for the geometric means MPN of *E. coli* of households' stored water quality. In Figure 9, the villages' geometric mean of households' stored water quality varies widely in the fourth, tenth, and thirteenth BWM rounds when measured by the inter-quartile range. The inter-quartile range of the first, third, fifth, seventh, ninth, eleventh, and fourteenth BWM rounds are tighter although outliers are more common in these rounds.

Figure 9: Geometric mean MPN of *E. coli* by BWM round



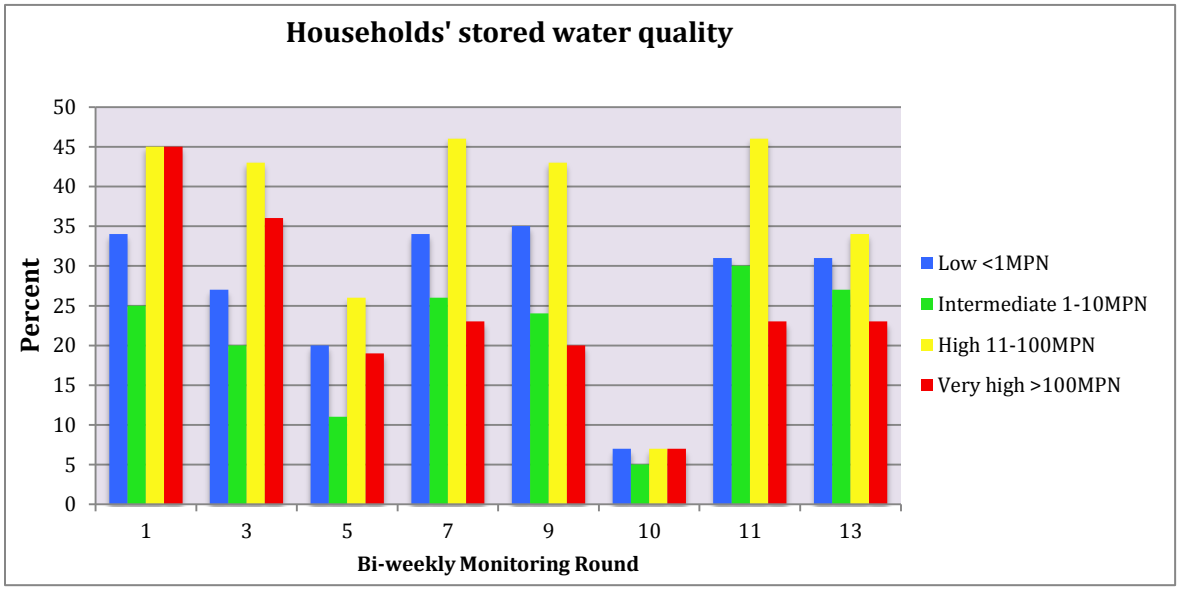
WHO drinking water risk categories by BWM round of BWM households and springs

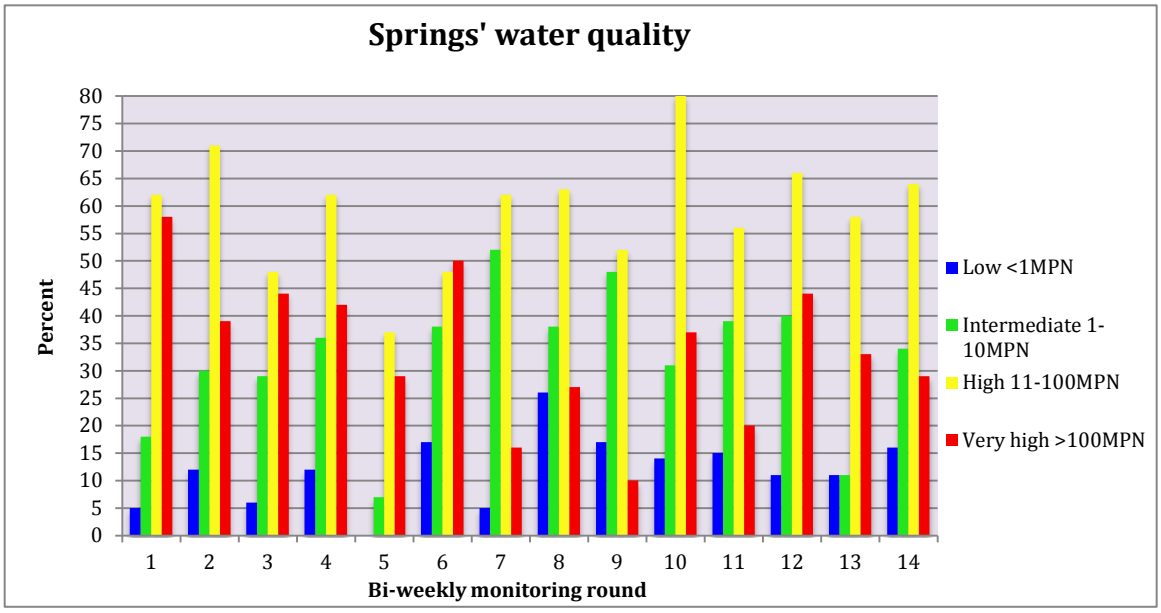
Water quality was examined using the ordered WHO risk levels for drinking water of households and springs (see Figure 10). It was hypothesized that the households'

stored water quality would not vary by BWM rounds. At the 95% confidence level, no statistically significant difference of households' stored water quality was detected using ordered WHO risk levels by the BWM rounds (Wald x^2=8.85, p=0.55).

It was hypothesized that the springs' water quality as measured by the WHO risk levels would not vary by BWM rounds. At the 95% confidence level, springs' water quality was not statistically different by ordered WHO risk levels by the BWM rounds (Wald x^2=15.13, p=0.30). In Figure 10, the high-risk category of 11-100 MPN of *E. coli* has the highest percentage of springs' water quality across all rounds except in round 6. Across the BWM rounds, springs' water quality in the low risk level ranged from 0 percent in round 5 to 25 percent in round 8.

Figure 10: WHO risk level by BWM rounds



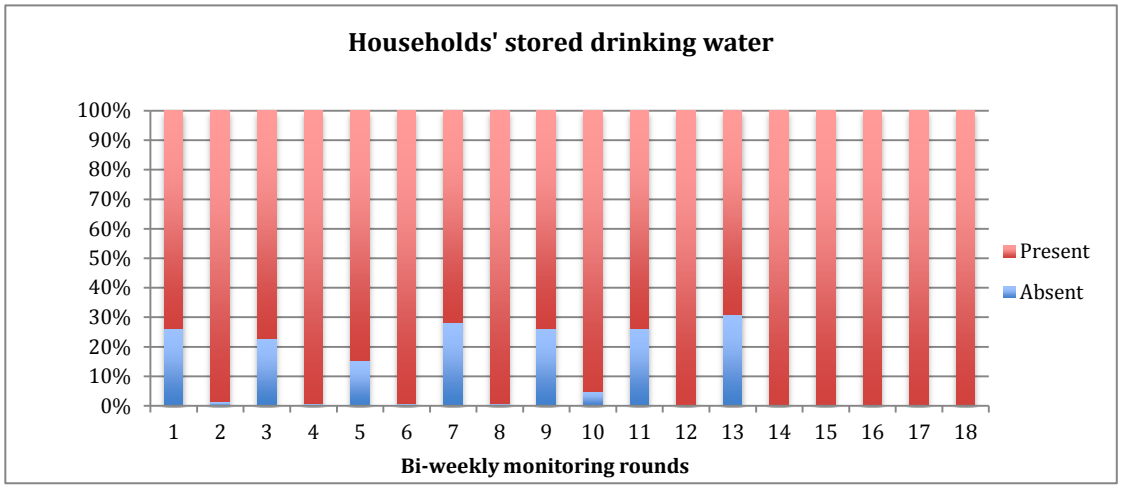


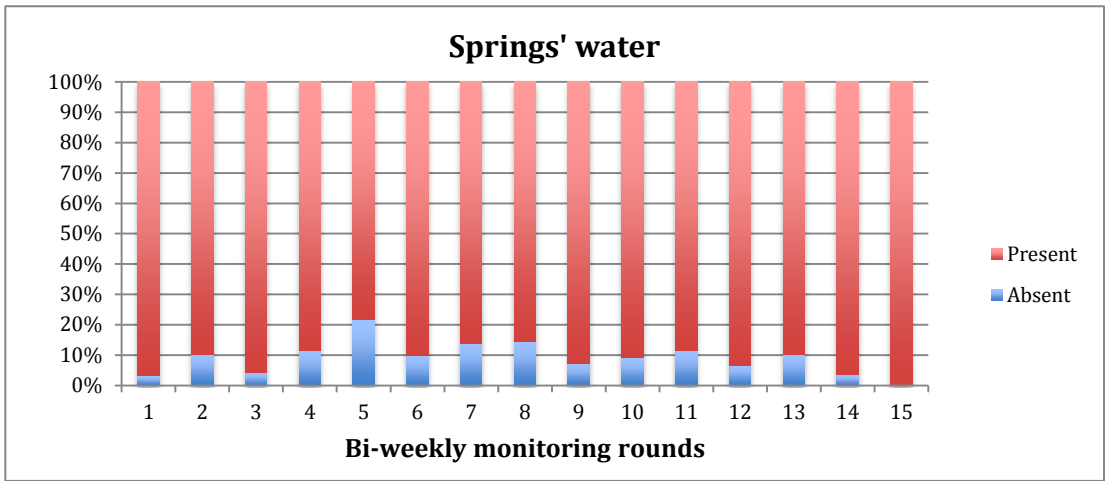
Presence absence test for E. coli by BWM round of BWM households and springs

It was hypothesized that the households' stored water quality would not be statistically different by BWM rounds using the presence absence test. At the 95% confidence level, the presence absence of *E. coli* test of households' stored water quality was not statistically different by the BWM rounds (Wald x^2=9.17, p=0.10). Households' stored water quality improved from springs' water quality in rounds 1, 3, 5, 7, 11, and 13. Households' stored water quality decreased from springs' water quality in rounds 6 and 8.

It was hypothesized that the springs' water quality would not be statistically different by BWM rounds using the presence absence test for *E. coli*. At the 95% confidence level, springs' water quality was statistically different by BWM rounds (Wald x^2=27.88, p=0.03). In Figure 11, at least 20% of springs in BWM rounds 6, 8 and 9 had no detectable levels of *E. coli.*

Figure 11: WHO risk level by BWM rounds





3.2 Subsample Overview

A summary of the statistical tests for the four methods of measuring water quality conducted on the subsample is found in Table 7. The baseline logarithmic MPN of *E. coli* and villages' geometric means of households' stored water quality F-statistic were not calculated. At the 95 percent significance level, the logarithmic MPN of *E. coli* detected a statistical difference across households' stored water quality at baseline and in the second survey round (see Table 7). At the 95 percent significance level, the ordered WHO risk levels method detected significant changes in households' stored water

quality at baseline. At the 95% significance level, there were not statistically significant changes across the ordered WHO risk levels of households' stored water quality in either the first, second, or third rounds. The villages' geometric means MPN of *E. coli* of households' stored water quality method detected significant differences at baseline ($F=2.90$, $p=0.1$) and in the first round ($F=4.15$, $p=0.002$) but not in the second or third rounds. The presence absence test also detected a significant difference in households' stored water quality at baseline and in round two at the 95 percent confidence level; significant differences in households' water quality across months of the year were not detected in either the first or third rounds using the presence absence test.

The logarithmic MPN of *E. coli* method detected significant differences of springs' water quality at the 95 percent confidence level at baseline. The ordinal WHO risk categories method also detected significant differences in springs' water quality at the 95 percent confidence level at baseline, but not in any subsequent rounds. At the 95 percent significance level, the presence absence test did not detect any significant differences in springs' water quality across the month of the year in any of the sub-sample's survey rounds.

Table 7: Sub-sample Statistics

Sub-sample		
	Households	**Springs**
Water Quality Comparison Method	**Month Statistical Test** for baseline, R1, R2, R3 (F or Wald x^2 statistic, p value)	**Month Statistical Test** for baseline, R1, R2, R3 (F or Wald x^2 statistic, p value)
Log MPN of E. coli	***Baseline:*** **F=6.03, p<0.0001** *Round 1:* F=1.07, p=0.38 ***Round 2:*** **F=8.73, p=0.0003** *Round 3:* F=1.22, p=0.30	***Baseline:*** **F=4.80, p=0.0001** *Round 1:* F=0.75, p=0.59 *Round 2:* F=0.82, p=0.52 *Round 3:* F=1.12, p=0.33
WHO risk levels	***Baseline*** **Wald x^2= 38.00, p<0.0001** *Round 1* Wald x^2= 7.28, p=0.28 *Round 2* Wald x^2= 0.71, p=0.49 *Round 3* Wald x^2= 1.12, p=0.57	***Baseline*** **Wald x^2= 28.16, p=0.0001** *Round 1* Wald x^2= 5.59, p=0.35 *Round 2* Wald x^2= 7.42, p=0.12 *Round 3* Wald x^2= 1.69, p<=0.43
Geometric means of MPN of E. coli	***Baseline:*** **F=2.90, p=0.01** ***Round 1:*** **F=4.15, p=0.002** *Round 2*: F=0.48, p=0.62 *Round 3*: F=0.84, p=0.96	Not possible
Presence absence test	***Baseline*****: Wald x^2= 14.41, p=0.03** *Round 1*: Wald x^2= 2.92, p=0.71 ***Round 2*****: Wald x^2= 38.01, p<0.0001** *Round 3*: Wald x^2= 1.14, p=0.57	*Baseline*: Wald x^2= 4.42, p=0.22 *Round 1*: Wald x^2 =5.29, p=0.07 *Round 2*: Wald x^2 =0.20, p=0.91 *Round 3*: Wald x^2 =4.81, p=0.09

The sample sizes for households and for springs in the sub-sample round by month are found in Table 8. In a round, months were excluded from the analysis if less than 11 households per month were sampled. In a round, months were excluded from the analysis if less than 10 springs per month were sampled.

Table 8: Sample size for the sub-sample by month of the year

Households												
Month	1	2	3	4	5	6	7	8	9	10	11	12
Baseline	131	30			1			40	283	262	196	230
Round 1				26	109	85	116	225	80			
Round 2								206	103	11	197	7
Round 3	99	109	27									
Total	230	139	27	26	110	85	116	471	466	273	393	237

Springs												
Month	1	2	3	4	5	6	7	8	9	10	11	12
Baseline	25	4			1			6	39	36	46	35
Round 1				5	22	22	26	42	13			
Round 2								42	21	2	34	2
Round 3	38	46	18									
Total	63	50	18	5	23	22	26	90	73	38	80	37

Seasonality as measured by Monthly Variability

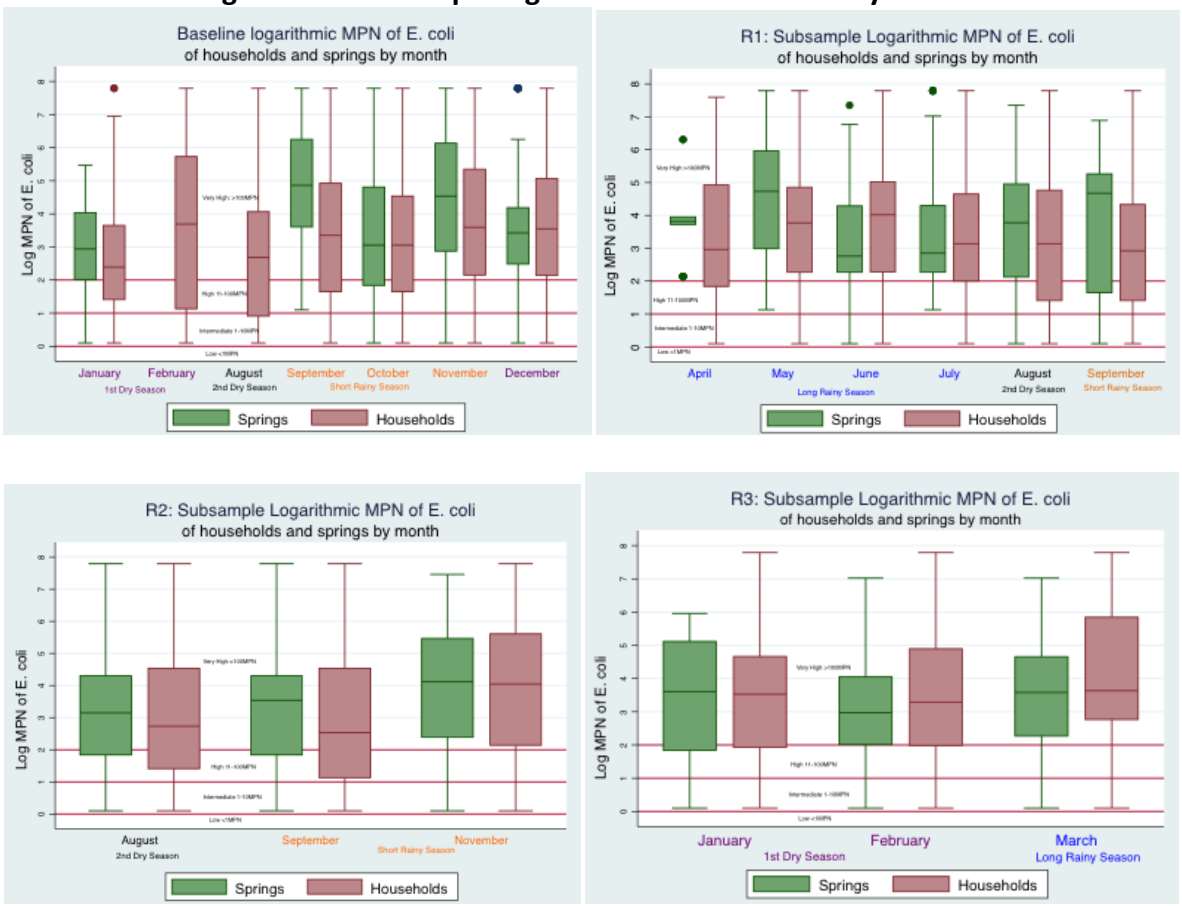
Logarithmic MPN of E. coli of Sub-sample's households and springs

In Figure 12, households' stored water quality and springs' water quality is displayed for each month of the year. It was hypothesized that the logarithmic MPN of *E. coli* of households' stored water quality would change across months within a year. A statistically significant change in water quality was detected using the logarithmic MPN of *E. coli* across months of the year at baseline ($F=6.03$, $p<0.0001$) and in the second round ($F=8.73$, $p=0.0003$) but not in the first or third rounds (see Table 7). The inter-quartile range for households' stored water quality has a wider variability that includes log levels below that of the springs' water quality inter-quartile range for baseline and for the second round. In the baseline survey, the inter-quartile range of households' stored water quality samples is above that of the springs' inter-quartile range; this pattern is reversed in the second round. Households' stored water quality is consistent across the months with log levels of MPN of *E. coli* varying between $10^{1.5}$ and 10^{6} in the first and third rounds.

It was hypothesized that the logarithmic MPN of *E. coli* of springs' water quality would change across months of the year. At the 95% confidence level, there was a significant difference of the logarithmic MPN of *E. coli* of springs' water quality by month of the

year at baseline (F=4.80, p=0.0001) but not in any of the sub-sample's subsequent rounds (see Table 7). Springs' water quality consistently varied across the rounds from 10^2 to 10^6 log levels of *E. coli*.

Figure 12: Sub-sample logarithmic MPN of *E. coli* by month

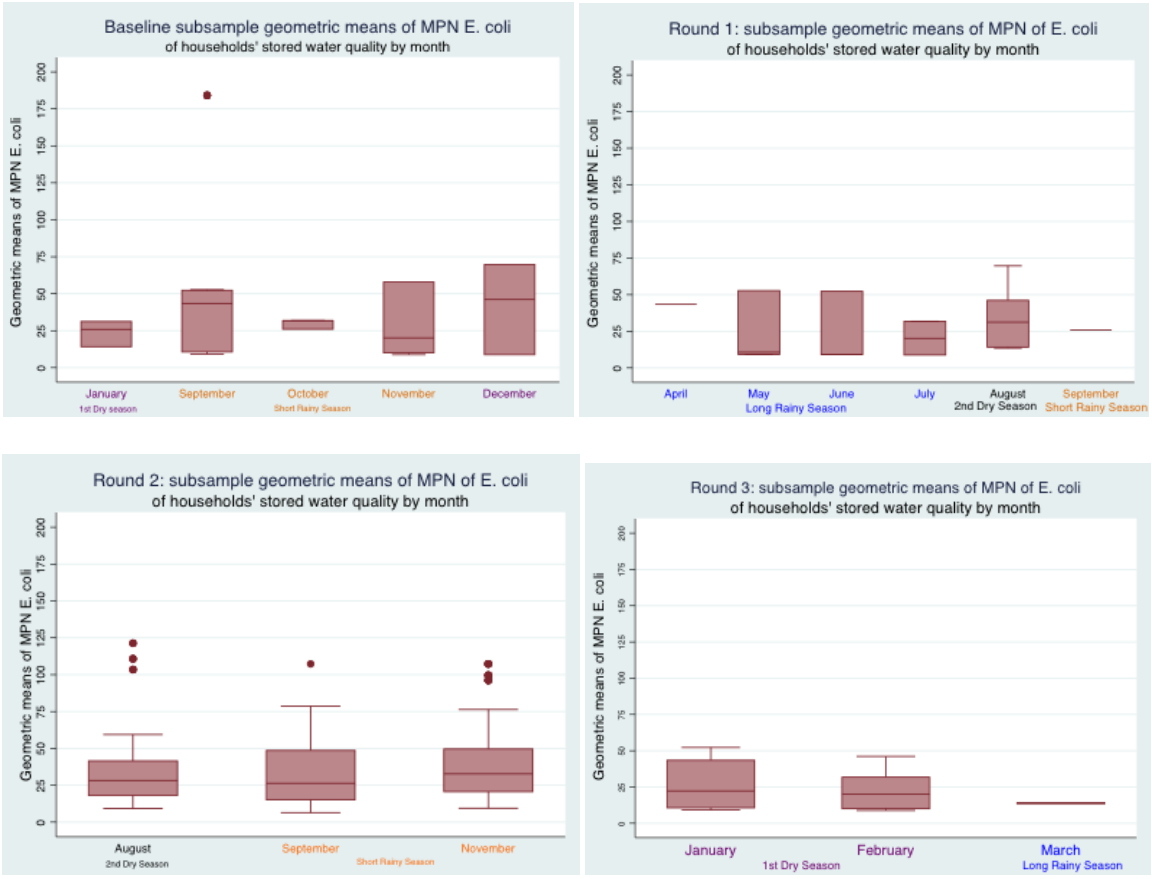


Geometric means of MPN of E. coli of sub-sample's households

It was hypothesized that the villages' geometric means MPN of *E. coli* of households' stored water quality would be different across the months of the year in each round (see Figure 13). When the clustered robust standard errors method of linear regression was performed at baseline, STATA returned a missing F-test statistic and p-value. At the 95% confidence level, the geometric means' MPN of *E. coli* method did not detect

significant differences of households' stored water quality in the second or in the third rounds (see Table 7). The geometric means method did detect a significant difference of households' stored water quality at baseline (F=2.90, p=0.01) and in the first round (F=4.15, p=0.002). The inter-quartile ranges of the geometric means of households' stored water quality are between 10 and 50 MPN of *E. coli* for the first through third rounds. At baseline, households' stored water quality samples for November and December have higher inter-quartile ranges of up to 75 MPN of *E. coli*.

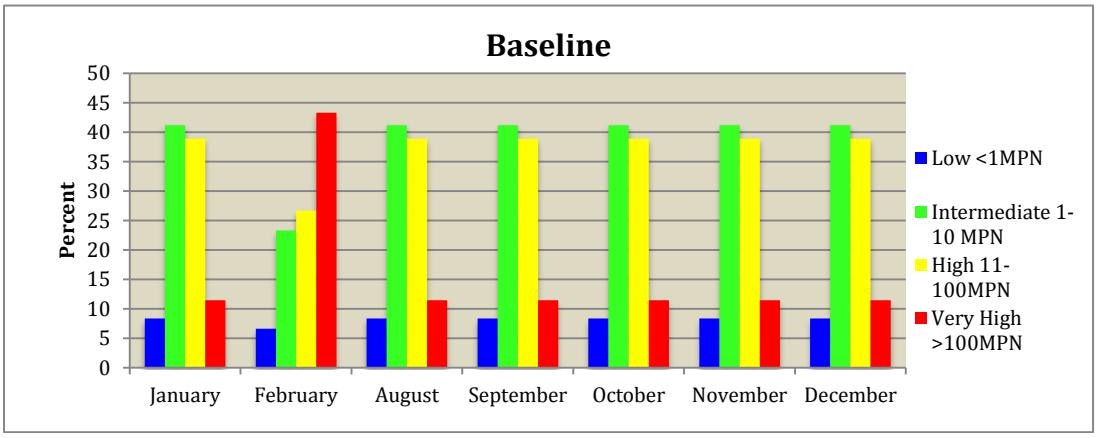
Figure 13: Sub-sample geometric means of MPN of *E. coli* by month

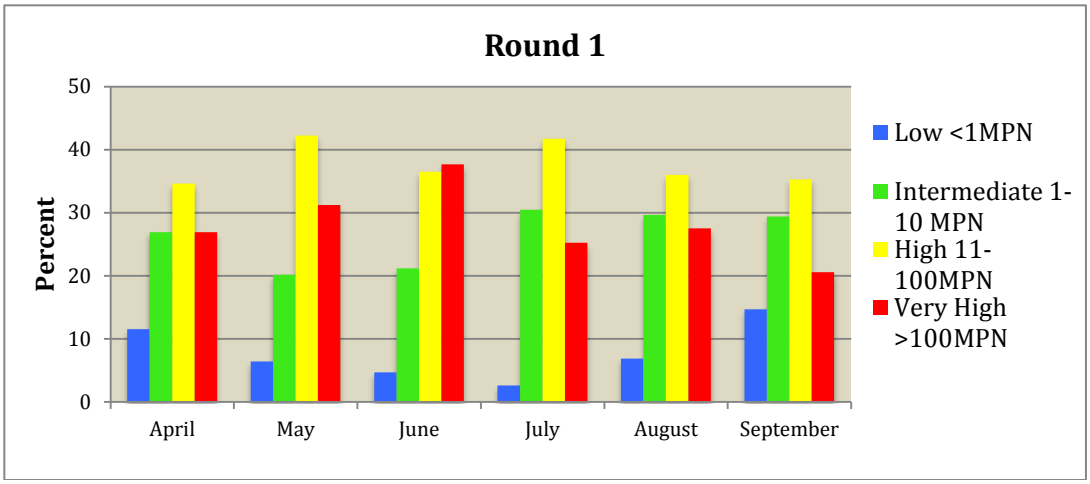


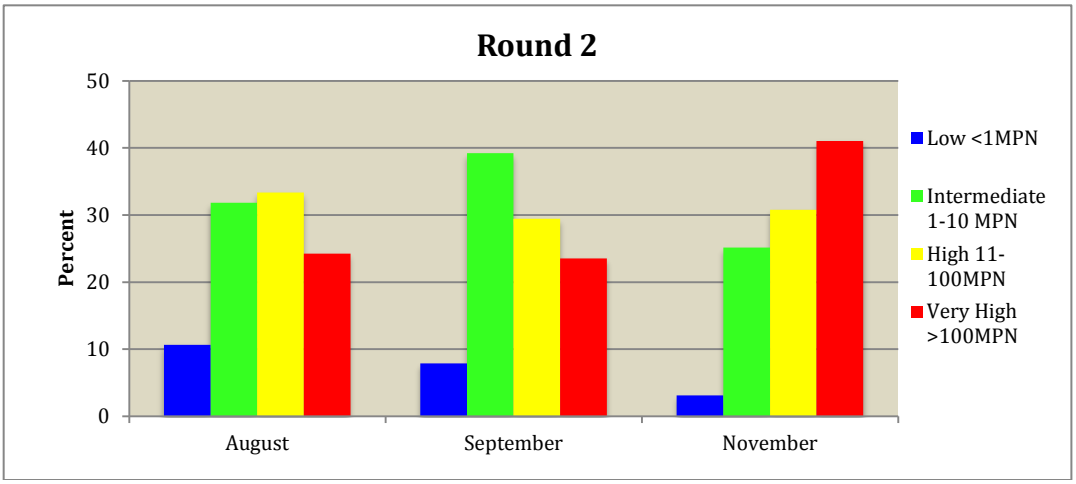
WHO drinking water risk categories of sub-sample households and springs

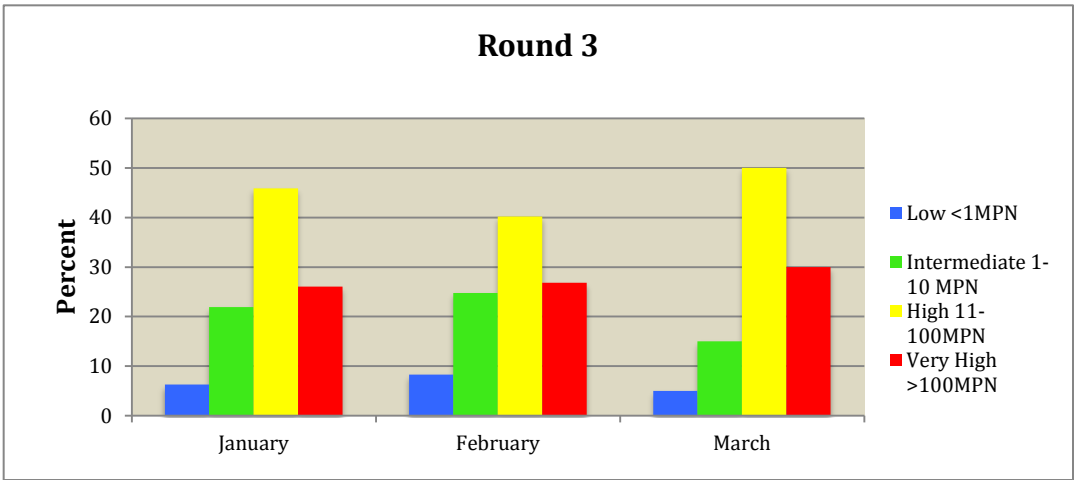
It was hypothesized that there would be a significant difference of households' stored water quality of the ordered WHO's drinking water risk categories across the months in the sub-sample (Figure 14). At the 95% confidence level, there was a detectable difference across the WHO's risk levels by month of the year for households' stored water quality at baseline (Wald x^2=38.00, p<0.0001). There was not a detectable difference across the WHO risk levels by month of the year for households' stored water quality in the first, second, or third rounds. In Figure 14, results from February (n=30) at baseline drive the statistical tests as all other months have similar percentages of households in each risk level category. At baseline, 7 percent (150 households) of all the households were in the WHO's low risk level, 27 percent (317 households) were in the intermediate and very high risk levels, and 38 percent (445 households) were in the high risk levels. In the first and third rounds, most households are in the high-risk level followed by the very-high risk level.

Figure 14: Sub-sample WHO risk levels of households' stored drinking water by month





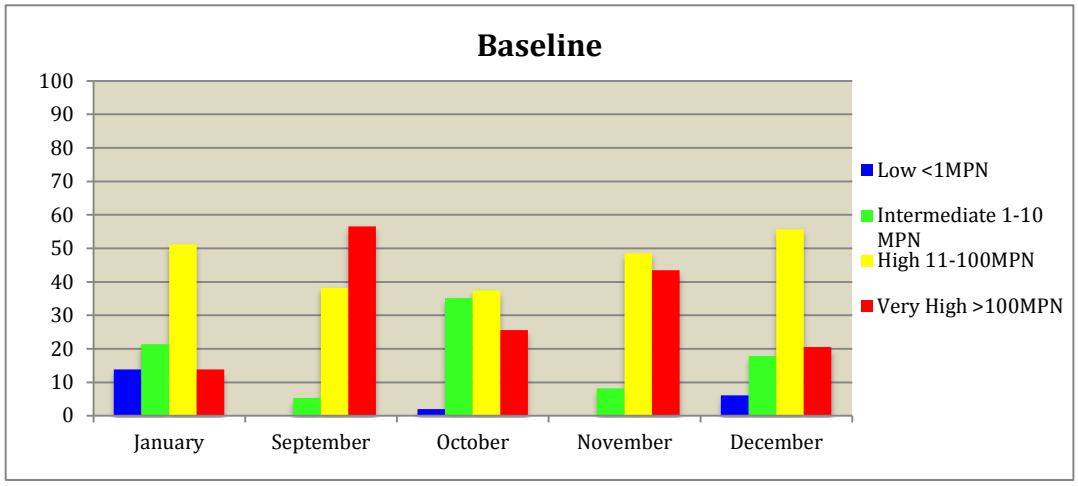


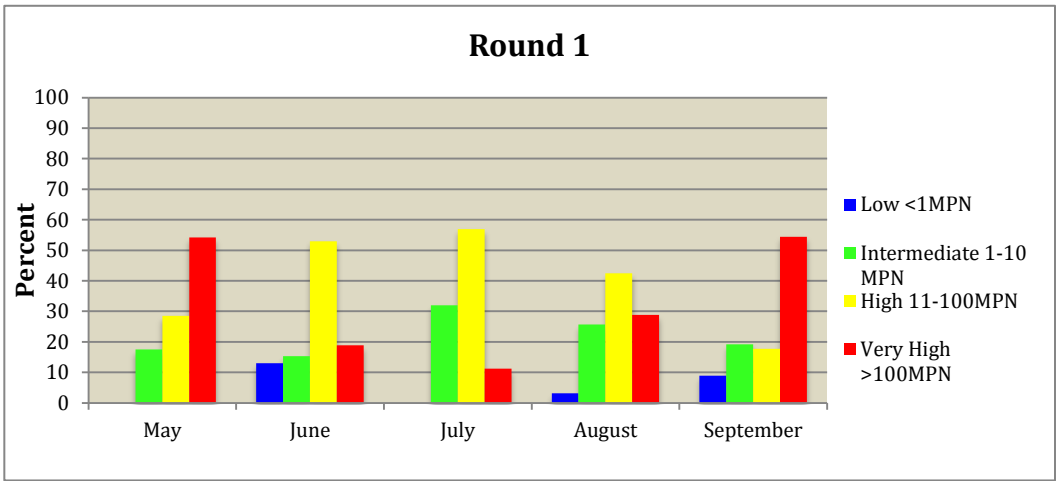


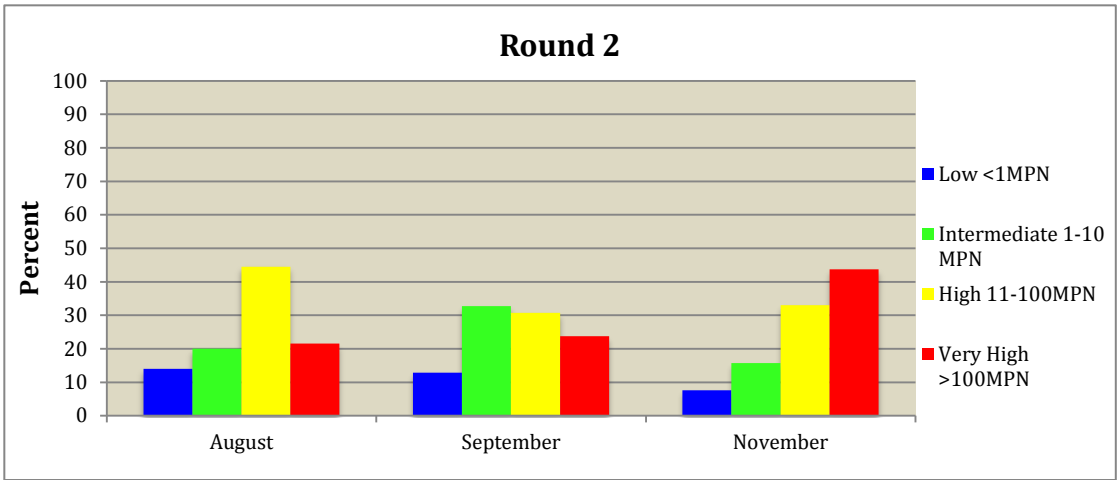
It was hypothesized that there would be a significant difference of springs' water quality of the ordered WHO's drinking water risk categories across the months of the year (Figure 15). At the 95 percent significance level, the ordered WHO risk categories

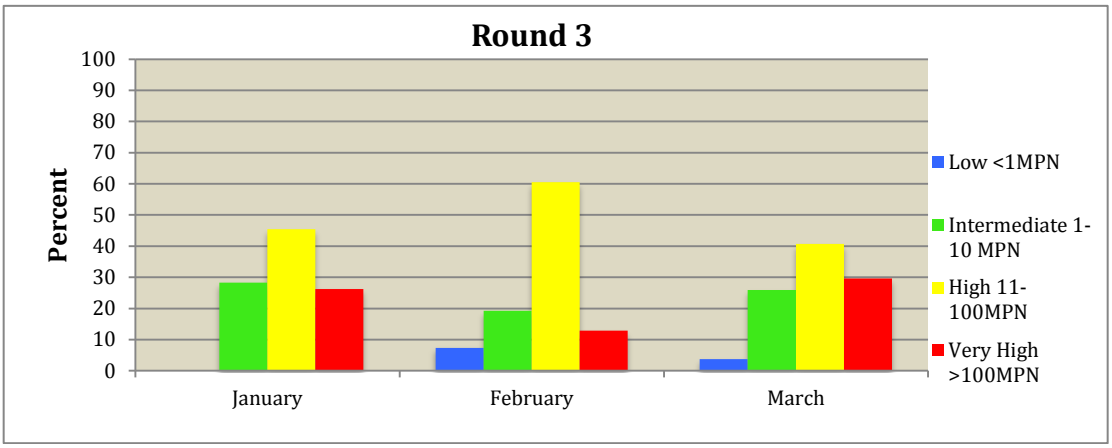
were statistically different across months of the year at baseline (Wald x^2=28.16, p=0.0001) but not in any of the subsequent rounds. At baseline, sample sizes range from 25 springs to 44 springs sampled per month. At baseline, 3 percent of springs were in the low risk category, 16 percent were in the intermediate category, 47 percent were in the high-risk category, and 34 percent were in the very-high risk category. In the first round, across the months of the year springs' water quality moved from high to very-high risk levels.

Figure 15: Subsample WHO risk levels of springs water by month





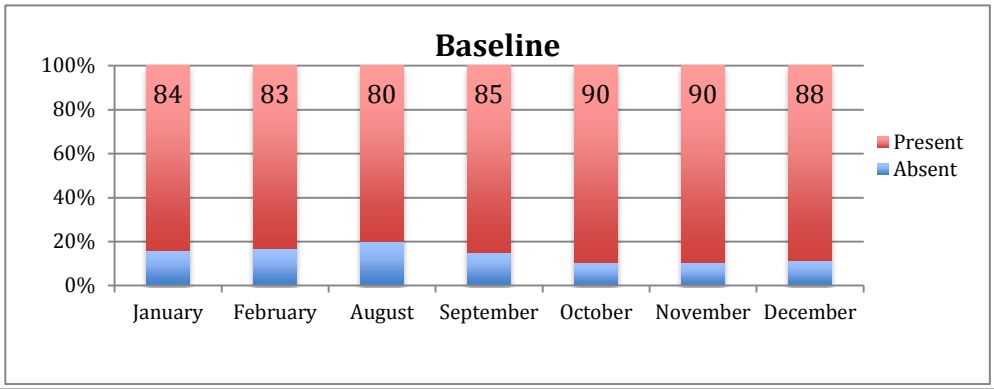


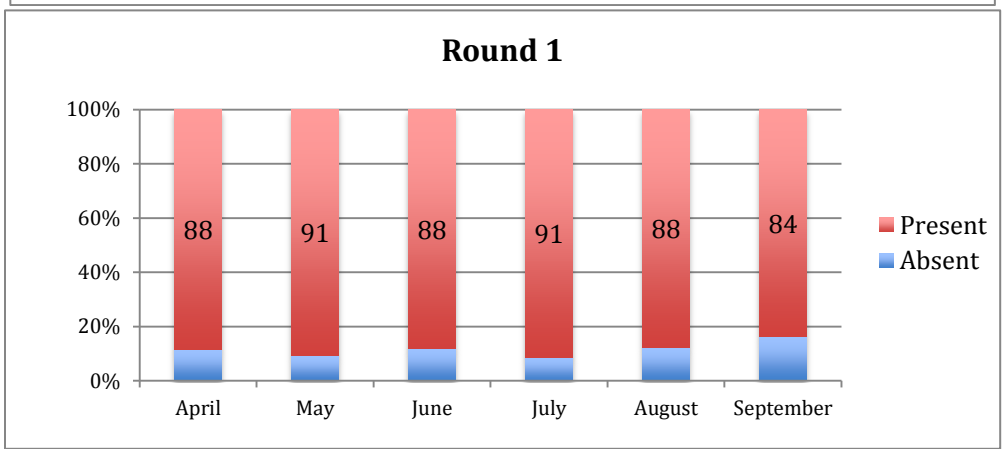


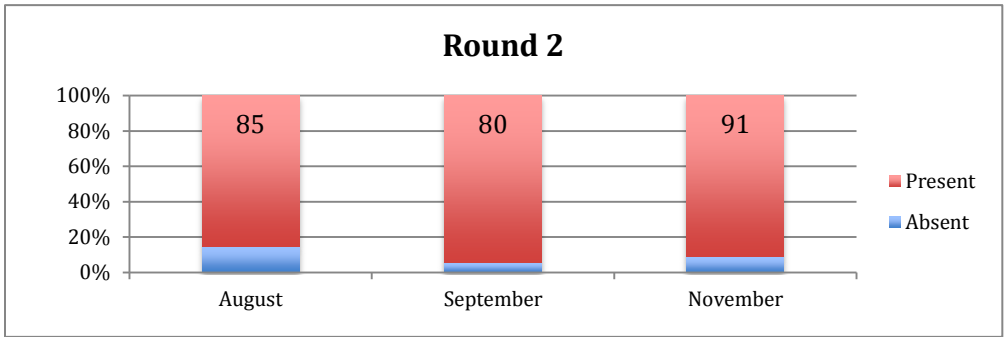
Presence Absence Test for E. coli of sub-sample's households and springs

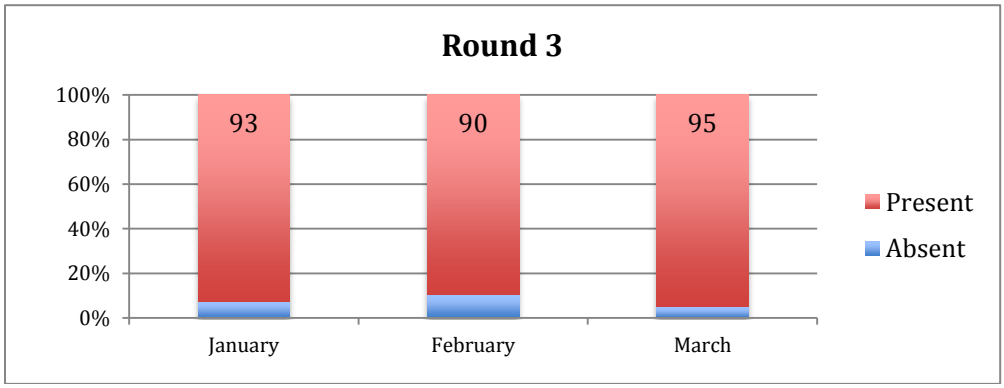
It was hypothesized that households' stored water quality would not be statistically different across the months of the year in the sub-sample's rounds (Figure 16). At the 95% confidence level, there was a significant difference detected between households' stored water quality across the months of the year at baseline (Wald x^2=14.41, p=0.03) and in the second round (Wald x^2=38.01, p<0.0001), but not in the first round or in the third round (see Table 7). Across all households' stored water quality, 12 percent at baseline, 11 percent in the first round, 14 percent in the second round, and 8 percent in the third round had no detectable levels of *E. coli*.

Figure 16: Sub-sample presence absence test of households by month



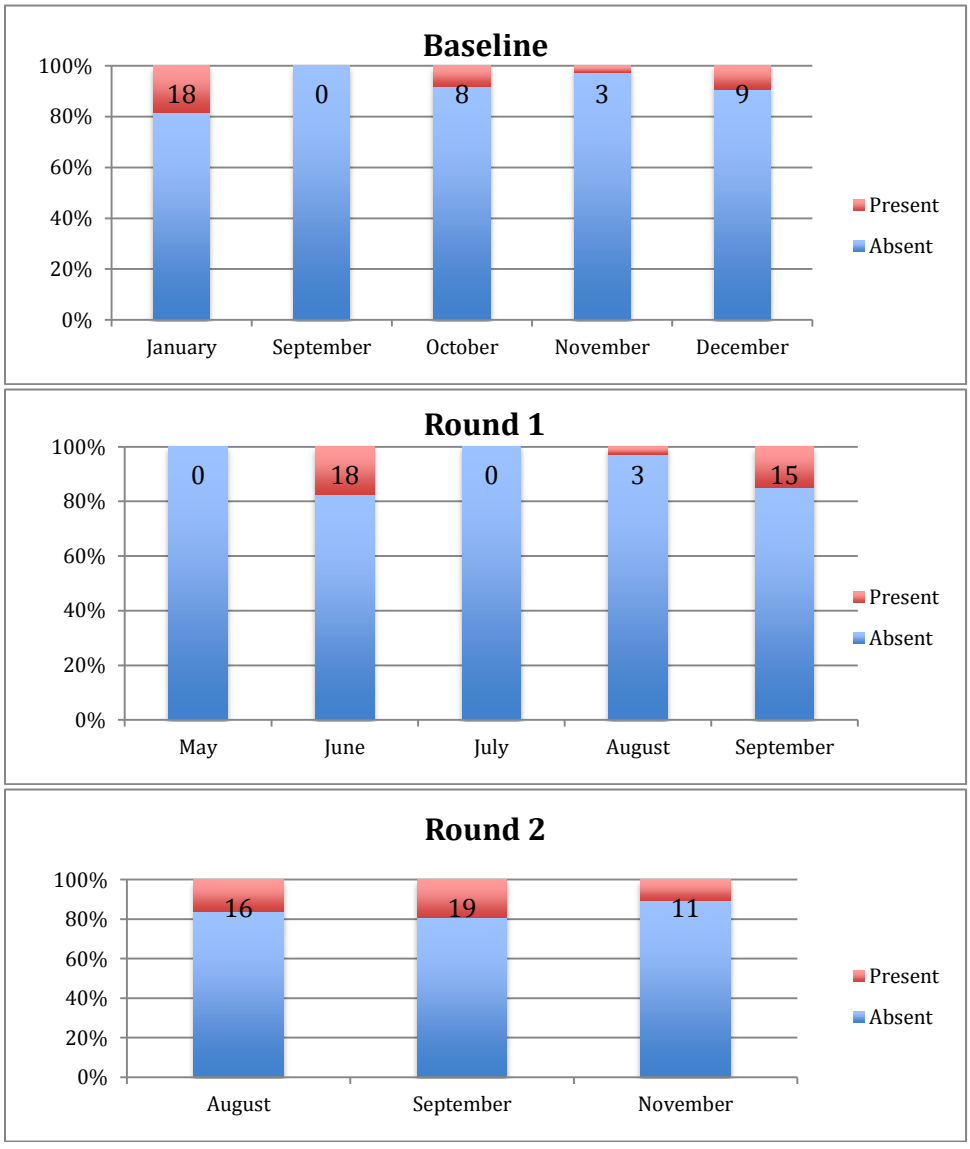


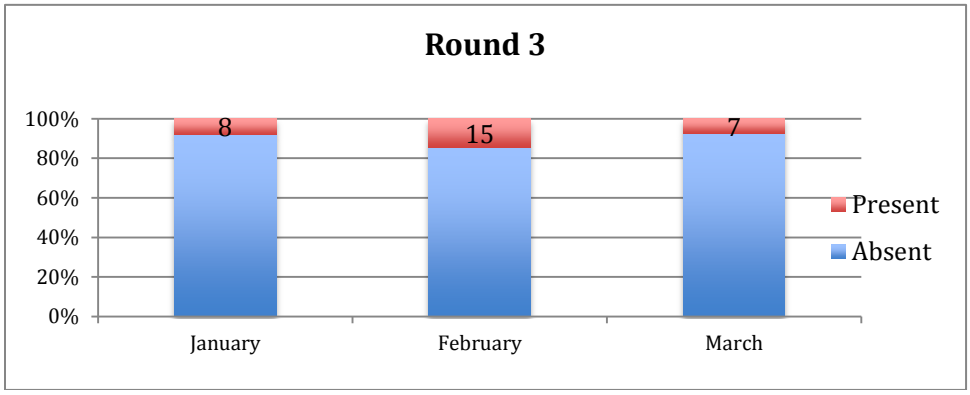




It was hypothesized that springs' water quality would not be statistically different across the months of the year in the sub-sample's rounds (Figure 17). Spring water quality has a higher percentage of *E. coli* that is absent from the samples. At the 95% confidence level, no significant differences were detected in springs' water quality across months of the year in any of the rounds (see Table 7). Spring water had a considerable number of months when *E. coli* was not detected from any of the spring samples. The sample size was large for some of these months. *E. coli* was not detected in September (n=283) at baseline, or in May (n=109) and in July (n=116) during the first round.

Figure 17: Subsample presence absence test of springs by month





3.3 Intra-cluster Correlation Coefficients (ICC)

Intra-cluster correlation coefficients for BWM and sub-sample rounds of households and springs are presented in Table 9. The ICCs of households' water quality is higher when using the logarithmic MPN of *E. coli* method (0.11) than when using the presence absence test for *E. coli* (0.02) in the BWM round. With an ICC of 0.11, 11% of the variability in households' stored water quality between clusters in the study can be described by the clustering effect of households using the same spring.

The ICCs of springs describe the similarity of water quality at the same spring across time. With an ICC of 0.46 of spring water, 46% of the variability of springs' water quality between springs in the study can be described by the clustering effect of springs' near location. The springs' ICCs in the BWM rounds are smaller than the ICCs of springs' samples across the sub-sample. The standard errors for both the springs and households ICCs of the logarithmic MPN of *E. coli* tests are slightly higher than the presence absence tests' standard errors in the BWM rounds. In the sub-sample, the standard errors of the ICCs are consistent across both methods.

The variance between clusters is higher for the logarithmic MPN of *E. coli* method than the presence absence test. Using either the logarithmic or the presence absence method, the variance within a cluster is always lower than the variance between clusters.

Table 9: Intra-cluster Correlation Coefficients

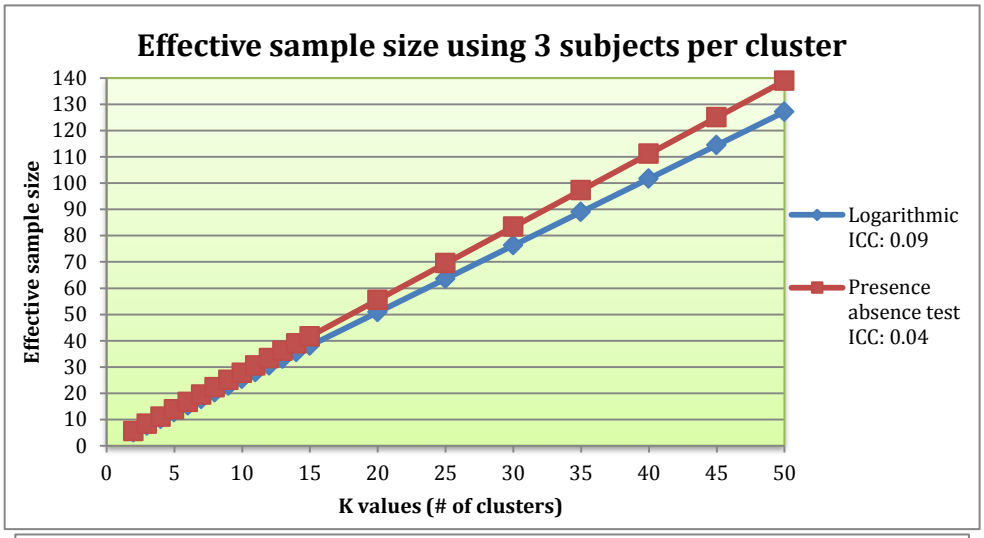
Bi-weekly monitoring samples of households' stored water quality				
Households	Household ICC Point (CI)	Standard Error	Between-cluster variance	Within-cluster variance
Logarithmic	0.09(0.001-0.18)	0.05	6.26	4.00
Presence absence	0.04 (0-0.12)	0.03	0.10	0.08
Springs	Spring ICC Point (CI)	Standard Error	Between-cluster variance	Within-cluster variance
Logarithmic	0.46 (0.33-0.58)	0.06	21.16	1.92
Presence absence	0.10 (0.02-0.18)	0.04	0.08	0.04
Sub-sample				
Baseline	Household ICC Point (CI)	Standard Error	Between-cluster variance	Within-cluster variance
Logarithmic	0.09 (0.04-0.14)	0.02	6.74	4.23
Presence absence	0.02 (0-0.06)	0.02	0.13	0.11
Round 1				
Logarithmic	0.09 (0.02-0.16)	0.04	5.69	3.78
Presence absence	0.008 (0-0.07)	0.03	0.10	0.10
Round 2				
Logarithmic	0.12 (0.04-0.20)	0.04	7.49	4.29
Presence absence	0.12 (0.04-0.20)	0.04	0.18	0.10
Round 3				
Logarithmic	0.01 (0-0.17)	0.08	3.96	3.81
Presence absence	0.11 (0-0.27)	0.08	0.09	0.07
Overall	Household ICC Point (CI)	Standard Error	Between-cluster variance	Within-cluster variance
Logarithmic	0.06 (0.03-0.09)	0.01	7.70	4.22
Presence absence	0.02 (0-0.04)	0.01	0.13	0.11
Overall	Spring ICC Point (CI)	Standard Error	Between-cluster variance	Within-cluster variance
Logarithmic	0.32 (0.21-0.42)	0.05	5.65	2.64
Presence absence	0.15 (0.04-0.26)	0.05	0.09	0.06

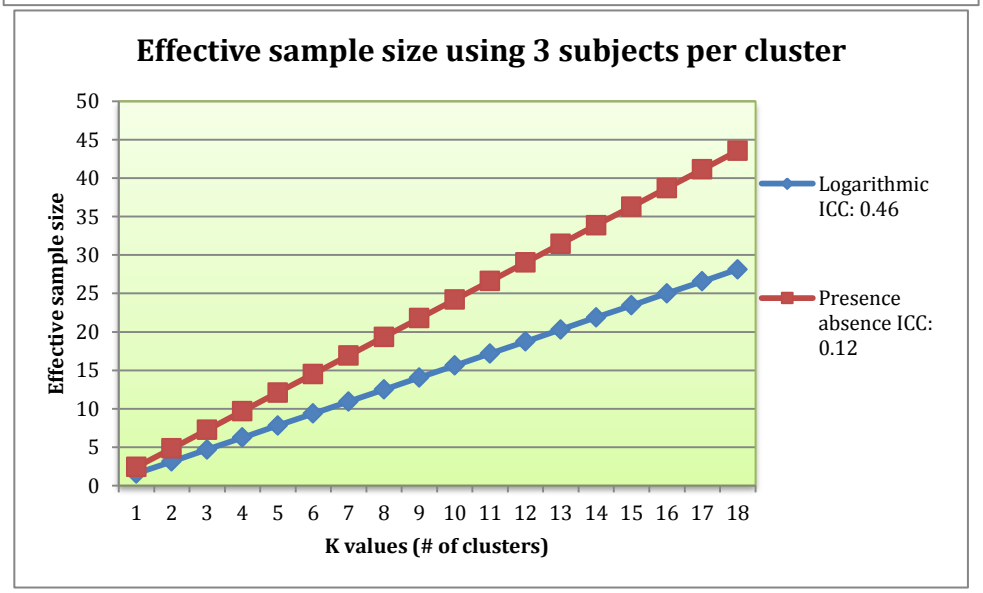
3.4 Power and Sample Size Calculations

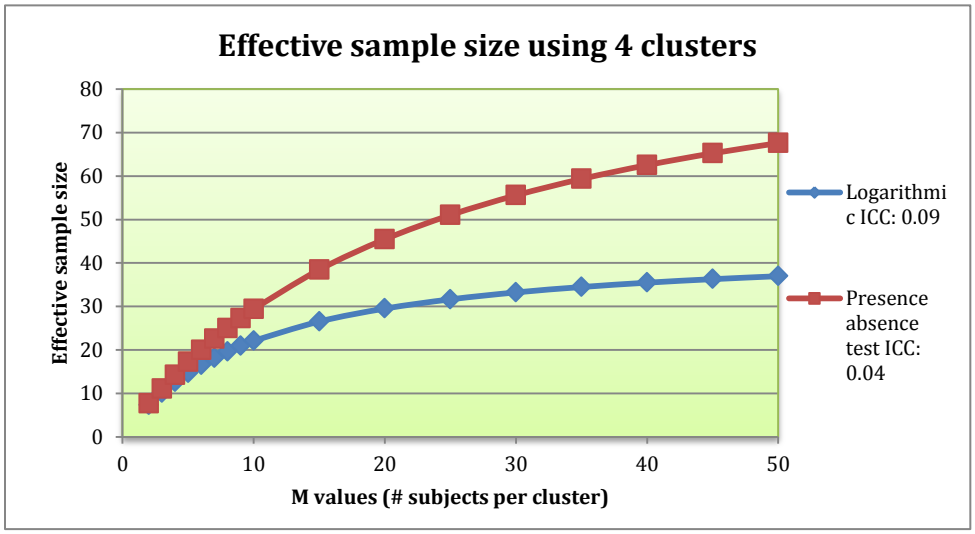
Using the ICCs of households' stored water quality and springs' water quality for the BWM round, the effective sample sizes were calculated using different number of clusters (k values) and different number of subjects in a cluster (m-values) for the presence absence test method and the logarithmic MPN of *E. coli* method (Figure 18). In each graph, the presence absence test method has a higher slope (springs: 2.3, households: 2.8) than the logarithmic MPN of *E. coli* method (springs: 1.7, households: 2.4). The effective sample size for the presence absence test method increases at a greater rate when increasing the number of samples per cluster compared to increasing the number of clusters sampled.

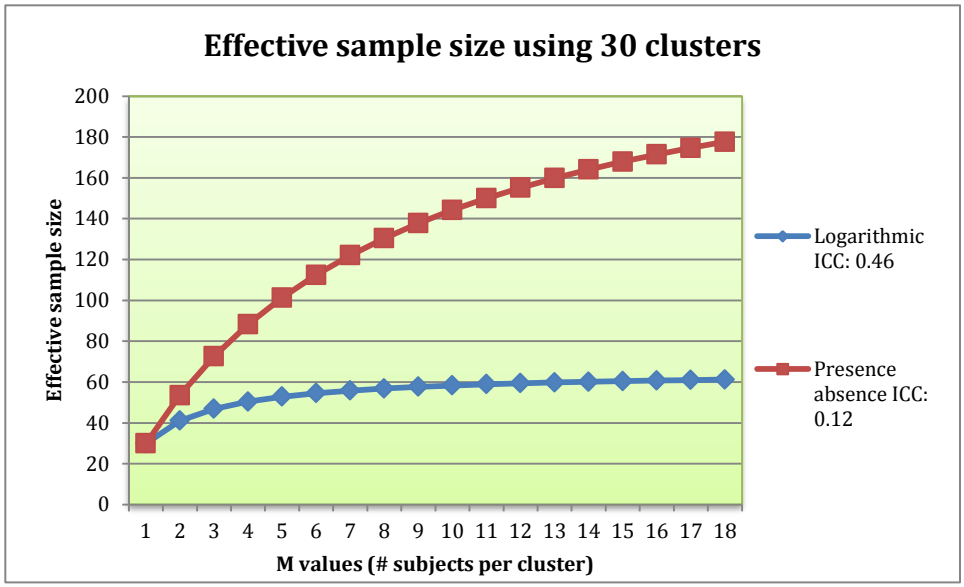
Springs' ICCs were higher than the ICCs of households' stored water. From 4 village clusters of 10 households, 22 households need to be sampled if using the logarithmic MPN of *E. coli* method and 29 households need to be sampled if using the presence absence test to be adequately powered. With 25 clusters of 3 households, the effective sample size for the logarithmic MPN of *E. coli* method is 64; the effective sample size for the presence absence test is 69 households. With 3 samples taken from 30 unprotected springs, the effective sample size is 47 using the logarithmic MPN of *E. coli* method and 73 using the presence absence method. If 14 samples are taken at 3 springs, then the effective sample size for the logarithmic MPN of *E. coli* method is 22 whereas the effective sample size for the presence absence method is 34.

Figure 18: Effective Sample Sizes








Effective sample size using 30 clusters
Effective sample size
200
180
160
140
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
M values (# subjects per cluster)
Logarithmic ICC: 0.46
Presence absence ICC: 0.12

4. Discussion

4.1 Households' stored water quality

It was expected that the BWM round with more than one observation per household would provide variability across the four water quality methods. The lack of significant results using all four water quality methods across the months of the year and across the BWM round for households' stored water quality was surprising. This indicates that households had consistently poor water quality across the rounds and across the months in the BWM sample. This implies that seasonality cannot be seen using households' stored water quality in a BWM round or across a few months of the year. Yet, this finding is not consistent with the water quality literature which recommends taking more than one sample from a household or a source over a given time period (Levy, Nelson et al. 2008; Levy, Hubbard et al. 2009). It is also possible that household water quality does not change widely in this population across the months of the year or in a bi-weekly sample.

The study was adequately powered for the BWM sample to detect changes in water quality using both the presence absence method and the logarithmic MPN of *E. coli* method for households. In the 15 bi-weekly monitoring rounds, 864 observations from 253 households in the BWM rounds were included in the analysis. Increasing the number of clusters is normally recommended over increasing the number of samples per cluster (Duflo, Glennerster et al. 2007). The ICC using the presence-absence test (0.04) is smaller than the logarithmic MPN of *E. coli* test (0.09) for households' stored

water quality. This means that a slightly larger number of samples are needed in order for the presence absence method to detect differences in water quality.

The sub-sample rounds provide another opportunity using a larger sample size to measure households' stored water quality. The presence absence test method and the logarithmic MPN of *E. coli* method both detected differences in households' stored water quality across the month of the year at the baseline survey and in the second round survey. The geometric means method detected differences of households' stored water quality across the month of the year at baseline and in the first round but not the in the second or third rounds. The WHO risk levels method only detected an ordered difference of households' water quality across the months of the year at baseline.

One of the driving goals of this project was to compare the presence absence test's capacity to detect differences in household water quality. The presence absence test gained from the using the Colilert method collapses statistically rich continuous data into a binary variable. Although the presence absence test method achieved the same results as the logarithmic method in both the BWM sample and the sub-sample, several limitations in the SIP-H study should be recognized before concluding that the logarithmic MPN of *E. coli* method and the presence absence tests are equal. The analysis of the sub-sample for each round did not consist of multiple water tests per household. The rounds could not be combined across all of the sub-sample because the water collection methodology changed and might have been a confounder in the analysis if all sub-sample rounds were combined.

Overall, this study provides support for both the presence absence tests and the logarithmic MPN of *E. coli* tests of households' water quality. Even though the presence absence test was calculated by reducing the variability shown through logarithmic MPN of *E. coli* method, each test yielded the same results in this case. It is expected that these results would be slightly different if the hydrogen sulfide method was used as it detects different microorganisms than the presence absence method used in this project. It is also expected that a larger number of samples would need to be taken using the presence absence test if an unbalanced design is used in the field. A larger number of samples per cluster will need to be taken to reliably detect variability using the presence absence test as compared to the smaller number of samples that could be taken with the logarithmic MPN of *E. coli* method. The number of samples per time unit taken should be balanced in future studies to determine if an effect is due to seasonality or to the study design. Due to the statistical calculations, the ordinal WHO risk categories method and the presence absence method may be more likely than the geometric means method and the logarithmic method to be influenced by an unbalanced design.

4.2 Springs' water quality

It was expected that springs' water quality would change in the BWM sample across both the BWM round and the month of the year. The three methods did detect differences across the BWM round time unit; however, only the presence absence test showed a significant difference in springs' water quality across the month of the year. It

was surprising that the presence absence test did not detect a difference in the springs' water quality across the months of the year in the sub-sample. Both the logarithmic MPN of *E. coli* method and the WHO risk levels method detected a significant difference in the second round of the sub-sample. The presence absence test does not have a higher statistical sensitivity than the logarithmic MPN of *E. coli* and WHO risk level methods in this study. The logarithmic MPN of *E. coli* was used to calculate the WHO risk levels and the presence absence measure. Intuitively, in this example, the differences detected by the presence absence test are artificial. The logarithmic MPN of *E. coli* enumeration method provides the most information as it maintains the continuity that other methods artificially divide at choice cut off points.

The difference between the presence absence test method's performance in the BWM sample and in the sub-sample could be explained by the different sampling strategies. Each spring was only sampled once across the months of the year in the sub-sample whereas each spring was sampled once in each of the 15 BWM rounds. This sampling methodology could also impact the variation within clusters and between clusters used to calculate the ICC. The ICCs of the households' stored water quality of the BWM rounds and sub-sample using the logarithmic MPN of *E. coli* method (BWM: 0.09, sub-sample: 0.06) and the presence absence test (BWM: 0.04, sub-sample: 0.02) were similar. The ICCs of the springs' water was higher for both the BWM rounds and the sub-sample using the logarithmic MPN of *E. coli* method (BWM: 0.46, sub-sample: 0.32) and the presence absence test (BWM: 0.10, sub-sample: 0.15).

In a research study designed to quantify the impact of improved water quality on health outcomes, the logarithmic MPN of *E. coli* method is still recommended to be used over the presence absence test.

4.3 Limitations

There are several limitations in the study. Water quality was not studied across the same months of the year in the sub-sample and in the BWM sample. The sampling strategy was not balanced across each month limiting the results. Therefore, seasonality of water quality across months was studied, but deciding which months of the year had reliably higher or lower water quality across the four years could not be stated. It was possible to compare months to each other within the BWM sample, but this was not the goal of the study. Other water quality studies that conclusively provide measurements and ranges of households' and springs' water quality for a given season based on samples from one season or from one year cannot reliably make these conclusions.

While the sample size within clusters of households' and springs' changed for this analysis in the sub-sample round, the large sample size accounted for the results. In the sub-sample, households' stored water was collected differently across rounds forcing a separate analysis for each round. Even though this analysis used data from a cluster randomized controlled trial, caution should be used when applying the ICCs of water quality calculated in western Kenya to other rural areas outside of Kenya.

5. Conclusion

5.1 Strengths

The SIP's cluster randomized controlled trial (RCT) sampling design provided a unique opportunity to examine the implications of the JMP's recommendations to use the presence absence test after 2015 to measure households' stored water quality. The frequency of water sample collection in the BWM sample is uses the gold standard that water quality research analysts desire. Households and spring protection in the sub-sample also uses the RCT design; yet the frequency of water sample collection mirrors the DHS' collection methods of water quality practitioners as one household water quality sample was taken per year. These two types of samples of water quality were used in summarizing the four main methods used in water quality research: the logarithmic MPN of *E. coli* method, the geometric means MPN of *E. coli* method, the WHO risk levels, and the presence absence test. Often, researchers and practitioners only present one method. Furthermore, the timely nature of this study provides necessary background to allow practitioners to prepare for changes in how water quality is collected.

5.2 Conclusions for Public Health Practitioners using the JMP recommendations

Public health practitioners will start using the JMP's recommendations for taking presence absence tests after 2015. Unless specified to return to households more than once, practitioners will take a one-time grab sample from households' stored water

quality in rural areas to fulfill these recommendations. If the time unit of analysis is month of the year, more household samples should be taken in months with the greatest variance in water quality. If the short rainy season is known to have a consistent level of household water quality, fewer samples should be collected from months in this season and a larger number of samples should be taken from months where the variance is unknown. If two samples of households' water quality will be taken for each household, it is better to collect all of the households' samples from a village in the same season. It is better to sample household water quality from 3 to 4 households in a village and to increase the number of villages sampled than to increase the number of households sampled within a cluster. Many more samples need to be taken using the presence absence method when compared to the logarithmic MPN of *E. coli* method. If the hydrogen sulfide presence absence test is used to perform this test, advantages include that it is easier to perform in the field, costs less, and does not require an incubator that would be necessary for conducting a presence absence test with logarithmic MPN of *E. coli* methods (Trottier 2010). Furthermore, the presence absence test allows practitioners to measure water quality in rural areas without the same time constraints of retreating to a field lab.

An overall percentage of the households that have safe water will be displayed for rural areas as the percentage of households' stored water quality without any detectable coliforms of *E. coli.* While these summary statistics are useful for donors, it is possible and recommended to detect changes of households' stored water quality across the months of the year in seasons of high interest. Since public health practitioners will

already be conducting the presence absence tests, it is possible to do so in a way that also allows for empirical evidence to identify specific target months for improving households' water quality.

5.3 Conclusions for Public Health Researchers

Public health researchers can use the intra-cluster correlation coefficients provided to add to the literature on the clustered nature of water quality. The ICCs from this study can reliably be used to calculate power for future studies in western Kenya. It is highly recommended that public health researchers report ICCs on the background variability of water quality as the opportunity to study household water quality increases. If the clustering effect of households that use the same source is ignored, researchers are likely to make very different conclusions about households' stored water quality and springs' water quality.

Researchers interested in the impacts of improved water quality rarely measure water quality after an intervention is implemented. Taking the short cut to measuring water quality as an exposure variable and diarrheal disease reduction as an outcome variable without a consistent water quality measurement across time fails the people that the intervention seeks to reach. The intermediate impact of water quality improvement should be measured given the multiple pathways of transmission of fecal pathogens and the analysis of these four water quality methods. The logarithmic MPN of *E. coli* test provides a high level of information and requires fewer samples for studies than the presence absence test.

The JMP's guidelines to conduct presence absence tests at the household level will be used to measure the risk of probable contamination of drinking water. If researchers use the logarithmic MPN of *E. coli* method in smaller studies and contribute in designing the JMP's sampling plans, researchers can assist practitioners by comparing results from more sensitive logarithmic MPN *of E. coli* method to presence absence tests for the study region.

References

Ahmed, N. U., M. F. Zeitlin, et al. (1993). "A longitudinal study of the impact of behavioural change intervention on cleanliness, diarrhoeal morbidity and growth of children in rural Bangladesh." Social Science and Medicine **37**(2): 159-171.

Ahuja, A., M. Kremer, et al. (2010). "Providing Safe Water: Evidence from Randomized Evaluations." 1-37.

Ashbolt, N. J., W. O. K. Grabow, et al. (2001). Indicators of microbial water quality. Water Quality: Guidelines, Standards and Health. L. Fewtrell and J. Bartram. London, IWA Publishing**:** 290-316.

Aziz, K. M., B. A. Hoque, et al. (1990). "Reduction in diarrhoeal diseases in children in rural Bangladesh by environmental and behavioural modifications." Transactions of the Royal Society of Tropical Medicine and Hygiene **843433-438**.

Bennett, H. B., A. Shantz, et al. (2010). "Characterisation of the water quality from open and rope-pump shallow wells in rural Cambodia." Water Sci Technol **61**(2): 473-479.

Blum, D., R. N. Emeh, et al. (1990). "The Imo State (Nigeria) Drinking Water Supply and Sanitation Project, 1. Description of the project, evaluation methods, and impact on intervening variables." Transactions of the Royal Society of Tropical Medicine and Hygiene **84**(2): 309-315.

Campbell, M. K., J. M. Grimshaw, et al. (2004). "Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported." BMC Medical Research Methodology **4**(9).

Cidavaenzi, M. T., M. Jere, et al. (1998). An evaluation of water urns to maintain domestic water quality. 24th WEDCC Conference. Islamabad, Pakistan.

Clasen, T., L. Roberts, et al. (2007). "Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis." **3**.

Duflo, E., R. Glennerster, et al. (2007). Using randomization in development economics research: a toolkit. London, Center for Economic Policy Research.

Dye, T. D. V., R. Apondi, et al. (2011). "A Qualitative Assessment of Beliefs, Attitudes, and Behaviors Related to Diarrhea and Water Filtration in Rural Kenya." American Journal of Public Health.

El Attar, L., A. A. Gawad, et al. (1982). "The sanitary condition of rural drinking water in a Nile delta village." The Journal of Hygiene **88**(1): 63-67.

Empereur-Bissonnet, P., V. Salzman, et al. (1992). "Evaluation of a new type of transport and storage recipient for improving the quality of drinking water." Bulletin de la Societe de Pathologie Exotique **85**: 390-394.

Esrey, S., J. P. Habicht, et al. (1986). "Infection, diarrhoea, and growth rates of young children following the installation of village water supplies in Lesotho." International Symposium on Water-related Health Issues: 11-16.

Esrey, S. A., J. B. Potash, et al. (1991). "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma." Bulletin of the World Health Organization **69**(5): 609-621.
Feachem, R. G., E. Burns, et al. (1978). Water, health and development. London, Tri-med books LTD.
Fewtrell, L., R. Kaufmann, et al. (2005). "Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis." Lancet **5**: 42-52.
Gasana, J., J. Morin, et al. (2002). "Impact of water supply and sanitation on diarrheal morbidity among young children in the socioeconomic and cultural context of Rwanda (Africa)." Environmental Research **90**(2): 76-88.
Gunther, I. and G. Fink (2011). Water and sanitation to reduce child mortality, The World Bank.
Heinanen, A., S. K. Chandiwana, et al. (1988). "Faecal contamination of rural drinking water in a commercial farming area in Zimbabwe." Central African Journal of Medicine **34**(11): 253-259.
Helsel, D. R. and R. M. Hirsch (1991). Statistical Methods in Water Resources, United States Geological Survey.
IDEXX. (2012). "A comparison of IDEXX Coliform and E. coli tests." Retrieved 1/11/12, 2012, from http://www.idexx.com/view/xhtml/en_us/water/comparison.jsf.
Iijima, Y., M. Karama, et al. (2001). "Prevention of bacterial diarrhea by pasteurization of drinking water in Kenya." Microbiology and Immunology **45**(6): 413-416.
Jalan, J. and E. Somanathan (2008). "The importance of being informed: experimental evidence on demand for environmental quality." Journal of Development Economics **87**: 14-28.
Jensen, B., J. H. Ensink, et al. (2003). "Effect of chlorination of drinking-water on water quality and childhood diarrhoea in a village in Pakistan." Journal of health, population, and nutrition **21**(1): 26-31.
Jensen, P. K., J. H. Ensink, et al. (2003). "Effect of chlorination of drinking-water on water quality and childhood diarrhoea in a village in Pakistan." J Health Popul Nutr **21**(1): 26-31.
JMP and GLAAS (2010). Second Meeting Strategic Advisory Group. New York, World Health Organization, UNICEF.
JMP Technical Task Force (2010). Meeting on monitoring drinking-water quality. Villie-Morgon, WHO, UNICEF**:** 1-46.
Joyce, T. M., K. G. McGuigan, et al. (1996). "Inactivation of fecal bacteria in drinking water by solar heating." Applied and Environmental Microbiology **62**(2): 399-402.
Kaltenthaler, E. C., B. S. Drasar, et al. (1996). "The use of microbiology in the study of hygiene behaviour." Microbios **88**(354): 35-43.
Kasirye, I. (2010). Household environmental conditions and disease prevalence in Uganda: the impact of access to safe water and improved sanitation on diarrhea, Centre for Environmental Economics and Policy in Africa.
Killip, S., Z. Mahfoud, et al. (2004). "What is an intracluster correlation coefficient? Crucial concepts for primary care researchers." Annals of Family Medicine **2**(3): 204-208.

Kirchhoff, L. V., K. E. McClelland, et al. (1985). "Feasibility and efficacy of in-home water chlorination in rural North-eastern Brazil." Journal of Hygiene **94**(2): 173-180.
Kremer, M., J. Leino, et al. (2007). Spring Cleaning: A Randomized Evaluation of Source Water Quality Improvement. Boston, Harvard University
University of California, Berkeley**:** 58.
Kremer, M., J. Leino, et al. (2011). "Spring Cleaning: rural water impacts, valuation, and property rights institutions " The Quarterly Journal of Economics **126**: 145-205.
Kremer, M., J. Leino, et al. (2011) "Spring cleaning: rural water impacts, valuation, and property rights institutions." The Quarterly Journal of Economics **126**.
Lehmsuluoto, P. O. (1986). Contamination of potable water in rural areas: should it be systematically monitored as part of impact management of various interventions. World Water '86**:** 41-42.
Levy, K., A. Hubbard, et al. (2009). "Drivers of Water Quality Variability in Northern Coastal Ecuador." Environmental Science and Technology(43): 1788-1797.
Levy, K., K. L. Nelson, et al. (2008). "Following the water: a controlled study of drinking water storage in northern coastal Ecuador." Environmental Health Perspectives **116**(11): 1533-1540.
Lindskog, R. U. and P. A. Lindskog (1988). "Bacteriological contamination of water in rural areas: an intervention study from Malawi." Journal of Tropical Medicine and Hygiene **91**(1): 1-7.
Luby, S., C. Mendoza, et al. (2008). "Difficulties in bringing point-of-use water treatment to scale in rural Guatemala." American Journal of Tropical Medicine and Hygiene **78**: 382-387.
Luby, S. P., M. Agboatwalla, et al. (2006). "Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial." Tropical Medicine and International Health **11**(4479-489).
Luoto, J. (2009). Information and persuasion: achieving safe water behaviors in Kenya.
Madajewicz, M., A. Pfaff, et al. (2007). "Can information alone change behavior? Response to arsenic contamination of groundwater in Bangladesh." Journal of Development Economics **84**: 731-754.
Mahfouz, A. A., M. Abdel-Moneim, et al. (1995). "Impact of chlorination of water in domestic storage tanks on childhood diarrhoea: a community trial in the rural areas of Saudi Arabia." Journal of Tropical Medicine and Hygiene **98**(2): 126-130.
Mazengia, E., M. T. Chidavaenzi, et al. (2002). "Effective and culturally acceptable water storage in Zimbabwe: maintaining the quality of water abstracted from upgraded family wells." Journal of Environmental Health **64**(8): 15-18, 25.
Mertens, T. E., M. A. Fernando, et al. (1990). "Childhood diarrhoea in Sri Lanka: a case-control study of the impact of improved water sources." Tropical Medicine and Parasitology **41**(1): 98-104.
Molbak, K., N. Hojlyng, et al. (1989). "Bacterial contamination of stored water and stored food: a potential source of diarrhoeal disease in West Africa." Epidemiology and Infection **102**(2): 309-316.
Morin, L., C. R. Jost, et al. (1990). Health benefits of water and sanitation in Rwanda. Water Quality Bulletin. UNICEF. **15:** 30-35.

Musa, H. A., P. Shears, et al. (1999). "Water quality and public health in northern Sudan: a study of rural and peri-urban communities." Journal of Applied Microbiology **87**: 676-682.
Nussbaumer, T. (2008). A study into the variation in microbiological water quality between different water source types. Water Management, Cranfield University. **Master of Science Thesis:** 1-51.
Office of Water (2002). Method 1103.1: Escherichia coli (E. coli) in water by membrane filtration using membrane-thermotolerant Escherichia coli Agar (mTEC), Environmental Protection Agency.
Olstadt, J., J. J. Schauer, et al. (2007). "A comparison of ten USEPA approved total coliform/E. coli tests." Journal of Water and Health **5**(2): 267-282.
Opiyo, P., W. Mukabana, et al. (2007). "An exploratory study of community factors relevant for participatory malaria control on Rusinga Island, western Kenya." Malaria Journal **6**(48): 1-20.
Pattanayak, S. K., C. Poulos, et al. (2007). "Informing the water and sanitation sector policy: case study of an impact evaluation study of water supply, sanitation and hygiene interventions in rural Maharashtra, India." RTI International: 1-30.
Petrifilm 3M. "3M Petrifilm E. coli Count Plate." Retrieved 2/26/12, 2012, from https://www.msu.edu/course/fsc/441/3mc&ec.html.
Pickering, A. J., J. Davis, et al. (2010). "Hands, water, and health: fecal contamination in Tanzanian communities with improved, non-networked water supplies." Environ Sci Technol **44**(9): 3267-3272.
Pinfold, J. V. (1990). "Faecal contamination of water and fingertip-rinses as a method for evaluating the effect of low-cost water supply and sanitation activities on faeco-oral disease transmission. I. A case study in rural north-east Thailand." Epidemiology Infectious **105**: 363-375.
Pinfold, J. V. and N. J. Horan (1996). "Measuring the effect of a hygiene behavior intervention by indicators of behaviour and diarrhoeal disease." Transactions of the Royal Society of Tropical Medicine and Hygiene **90**(4): 366-371.
Platenburg, R. J. P. M. and M. Zaki (1993). "Patterns of water quality in rural areas of Assyut governorate, Egypt." Water science and technology **27**(9): 55-65.
Rajasekaran, P., P. R. Dutt, et al. (1977). "Impact of water supply on the incidence of diarrhoea and shigellosis among children in rural communities in Madurai." Indian Journal of Medical Research **66**(2): 189-199.
Sandiford, P., A. C. Gorter, et al. (1989). "Determinants of Drinking Water Quality in Rural Nicaragua." Epidemiology and Infection **102**(3): 429-438.
Schmidt, W. P. and S. Cairncross (2009). "Household water treatment in poor populations: is there enough evidence for scaling up now?" Environmental Science and Technology **43**(4): 986-992.
Shears, P., M. A. Hussein, et al. (1995). "Water sources and environmental transmission of multiply resistant enteric bacteria in rural Bangladesh." Annals of Tropical Medicine and Parasitology **89**(3): 297-303.
Shiffman, M. A., R. Schneider, et al. (1978). "Field studies on water sanitation and health education in relation to health status in central America." Progress in Water Technology **11**: 143-150.

Simango, C., J. Dindiwe, et al. (1992). "Bacterial contamination of food and household stored drinking water in a farmworker community in Zimbabwe." Central African Journal of Medicine **38**(4): 143-149.
Smeeth, L. and E. Siu-Woon Ng (2002). "Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC trial of the assessment and management of older people in the community." Controlled Clinical Trials **23**: 409-421.
Statistical Consulting Group. "Analyzing correlated (clustered) data." Retrieved 4/14/12, 2012, from http://www.ats.ucla.edu/stat/stata/library/cpsu.htm.
Statistical Consulting Group. "Stata Data Analysis Examples Logistic Regression." Retrieved 4/14/12, 2012, from http://www.ats.ucla.edu/stat/stata/dae/logit.htm.
Stauber, C. E., G. M. Oritz, et al. (2009). "A randomized controlled trial of the concrete biosand filter and its impact on diarrheal disease in Bonao, Dominican Republic." American Journal of Tropical Medicine and Hygiene **80**(2): 286-293.
Sutton, S. and D. Mubiana (1989). "Experience from a water quality project in Zambia." Waterlines **8**: 20-21.
Tiwari, S. S., W. P. Schmidt, et al. (2009). "Intermittent slow sand filtration for preventing diarrhoea among children in Kenyan households using unimproved water sources: randomized controlled trial." Tropical Medicine and International Health **14**(11): 1374-1382.
Tomkins, A. M. and B. S. Drasar (1978). "Water supply and nutritional status in rural Northern Nigeria." Transactions of the Royal Society of Tropical Medicine and Hygiene **72**(3): 239-243.
Trevett, A. F., R. C. Carter, et al. (2004). "Water quality deterioration: A study of household drinking water quality in rural Honduras." International Journal of Environmental Health Research **14**(4): 273-283.
Trottier, S. (2010). Study of four new, filed-based, microbiological tests: verification of the hydrogen sulfide (H2S), Easygel, Colilert and Petrifilm tests. Civil and Environmental Engineering, Massachusetts Institute of Technology. **Master of Engineering in Civil and Environmental Engineering**.
UNICEF and World Health Organization (2009). Diarrhoea: why children are still dying and what can be done, The United Nations Children's Fund and the World Health Organization**:** 1-68.
USGS. (2007). "Colilert method for total coliforms and Escherchia coli." Retrieved 1/11/12, from http://oh.water.usgs.gov/.
van der Hoek, W., F. Konradsen, et al. (2001). "Irrigation water as a source of drinking water: is safe use possible?" Tropical Medicine and International Health **6**(1): 46-54.
Verweij, P. E., M. van Egmond, et al. (1991). "Hygiene, skin infections and types of water supply in Venda, South Africa." Transactions of the Royal Society of Tropical Medicine and Hygiene **85**(5): 681-684.
Waddington, H., B. Snilstveit, et al. (2009). International initiative for impact evaluation, 3ie**:** 1-116.

Wang, Z. S., D. S. Shepard, et al. (1989). "Reduction of enteric infectious disease in rural China by providing deep-well tap water." Bulletin of the World Health Organization **67**(2): 171-180.
World Health Organization (2011) "Guidelines for Drinking-water Quality."
World Health Organization. (2012). "Burden of disease and cost-effectiveness estimates." Retrieved 1/7/12, 2012, from http://www.who.int/water_sanitation_health/diseases/burden/en/.
World Health Organization and UNICEF (2006) "Meeting the MDG drinking water and sanitation target the urban and rural challenge of the decade."
World Health Organization and UNICEF (2010). Progress on sanitation and drinking-water. 2010 update.
Wright, J., S. Gundry, et al. (2004). "Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use." Tropical Medicine and International Health **9**(1): 106-117.

Appendix

(Jensen, Ensink et al. 2003)

The effect of chlorination on drinking water quality and on childhood diarrhea was studied in two Pakistani villages. The link between bacteriological safe water and diarrheal reduction was examined. In these villages, water was pretreated with slow sand filtration before it was pumped twice a day through irrigation canals to homes (Jensen, Ensink et al. 2003). In the intervention village, chlorine was added to the water before it reached the village. *E. coli* medium m-Coli Blue 24 tests of source water quality at three points were collected once daily and thrice weekly for a year. At baseline before chlorination, the intervention village had statistically significant better water quality (13.3 per 100 ml) as measured by geometric means of *E. coli* than the control village (137.0 per 100 ml) (Jensen, Ensink et al. 2003). The geometric means of water quality decreased in both villages over time (3 per 100ml in the intervention village; and 49 100mL in the control village) (Jensen, Ensink et al. 2003). Because baseline water quality and treatment characteristics between villages were dissimilar, researchers could not conclude that chlorination improved water quality. Both villages also treated their water via sedimentation and filtration. Protecting water via chlorination was statistically significant and *negatively* associated with having a higher incidence of diarrhea (8.7 vs. 3.6 episodes per 103 person days) in the intervention village (Jensen, Ensink et al. 2003). This small study of two dissimilar villages used 53 water tests throughout the year: 20 for the intervention village and 30 for the control village.

(Joyce, McGuigan et al. 1996)

The SODIS method was tested in the rift valley of northwestern Kenya. Water from a local river source, used by Esonorua village, was collected at 8 am on each sampling day for the tests (Joyce, McGuigan et al. 1996). Temperature, turbidity (NTUs), and solar power measurements were taken hourly; on average there was 8600 colony forming units per milliliter of *E. coli* (Joyce, McGuigan et al. 1996).

Only 20 tests were conducted with 15 tests in August (short rainy season) and 5 tests in February (first dry season). Of these tests, coliform units fell sharply from 10^7 to below 10^1 log levels of contamination after four to seven hours if the temperature of the air was above 45.6°C; bottles tested with air temperature of 36.3°C showed bacterial re-growth after four and eight hours (Joyce, McGuigan et al. 1996).

(Kirchhoff, McClelland et al. 1985)

The link between improved water quality via point of use chlorination and diarrheal cases in 20 households of Sao Joao was investigated. High diarrheal rates coincided with heavy rainfall at the beginning months of the year when *Enterotoxigenic E. coli* samples were isolated from drinking water samples (Kirchhoff, McClelland et al. 1985). A community health worker visited each household and deposited 1 milliliter of either placebo or hypochlorite solution into the drinking water for 9 weeks. After the second nine weeks, placebo and intervention households were switched. Three household water samples and source samples were collected with one sample taken before and one sample taken after the switch. Medical students interviewed mothers three times a

week about their children's diarrhea. In the intervention households, no significant decrease of diarrheal cases occurred (p=0.48) (Kirchhoff, McClelland et al. 1985). In both the control and the intervention groups, diarrheal cases decreased after the switchover.

Kirchoff and colleagues did not specify how often the community health worker visited the households. Visiting once a week is quite different than visiting every day and can account for why no difference was seen in reduced diarrheal cases. Researchers did acknowledge that the organic load, water accessed outside the household, and using a chlorine dosage below the WHO recommended standard could have affected the results of the trial. In households given hypochlorite, 5 to 9 year old children were reported to have increased diarrheal rates during a 12 day period (p=0.04) (Kirchhoff, McClelland et al. 1985). Since it is not clear when only the 3 household water samples were collected, even if a difference in diarrheal cases had been detected it would have been a stretch to state that this difference was due to improved water quality.

(Iijima, Karama et al. 2001)

Point of use pasteurization was tested to prevent severe diarrheal cases in four villages in Malindi, Kenya. Villagers generally used turbid pond water but also had access to river water; pooled rainwater was used during the rainy season. Water was tested for *E. coli, salmonella, S. dysenteriae,* and *V. cholerae.* Household water was collected from households were the pasteurization method was shown 6 times within 3 months. Each of the 1,500 households was visited every two to three weeks for four months and

surveyed about diarrhea cases. Household water quality improved over time with 10.7 percent of households to 43.1 percent of households demonstrating non-detectable levels of *E. coli* ($p<0.0001$, $x^2=69.67$) (Iijima, Karama et al. 2001). The odds of cases of severe diarrhea among households that pasteurized their water was significantly lower than the odds of cases of severe diarrhea among households that did not pasteurize their water (OR= 0.55, p=0.002) (Iijima, Karama et al. 2001).

It is not clear how many of the 1,500 households were shown how to pasteurize their water six times. Therefore, it is hard to understand the context of the increase in non-detectable levels of *E. coli* of household water. If only 50 of the 1,500 households completed six pasteurization observations this would be different than the entire 1,500-targeted households pasteurizing their water. Diarrhea measurements may have been biased by the Hawthorne effect.

(Mahfouz, Abdel-Moneim et al. 1995)

In nine villages of Saudi Arabia's Ashir region, 220 wells were tested for chemical and microbiological contamination. A voluntary chlorination study was conducted with a total of 325 households. Households' tanks were tested the day after chlorination was done for residual chlorine. Stool samples were taken at baseline and six months later. 92% of the wells were positive for *E. coli* and all had coliforms; 9% of the wells exceeded the sulfate and hardness levels recommended by the WHO (Mahfouz, Abdel-Moneim et al. 1995). Less than 4 percent of households using the chlorine tested positive for bacteriological contamination but all random samples taken from

households were positive for coliforms and for *E. coli* (Mahfouz, Abdel-Moneim et al. 1995). The presence of *giardia, E. histolytica*, and *E. coli* in children's stools decreased significantly in households that chlorinated their water after six months. Due to the drop out of the control group, results could not be compared across time which is why it is not possible to attribute these lowered levels to improved water quality.

(Mertens, Fernando et al. 1990)

Although Mertens et al conducted a retrospective case-control study, they reported on household water quality over 40 weeks in different households. Due to their retrospective study design, they did not find an association of diarrheal cases and water quality. In the rainy season, 25 milliliters of water was filtered and analyzed whereas in the dry season only 5 milliliters of water was filtered and analyzed. Households had better water quality if they boiled their water, stored it in a non-earthenware container, covered the container, and used different containers for boiling and for storing water. Contamination of household water peaked in the beginning of the rainy season and the middle of the dry season(Mertens, Fernando et al. 1990).

(Wang, Shepard et al. 1989)

In 12 villages of Qidong County, China, a prospective case-control study of deep-well tap water and enteric infectious disease was conducted. Six of the villages had been using deep-well tap water for two to three years whereas six control villages consumed surface water. Average water quality from the control villages was 772 colony-forming units per liter whereas the average deep-well tap water quality was 2.3 colony-forming

units per liter. Wang and colleagues concluded that deep-well tap water reduced incidence rates of viral hepatitis A, cholera, and acute watery diarrhea (Wang, Shepard et al. 1989).

Water collection methods were not presented. This is an important issue as the timing and sample size are necessary to determine study validity. For example, the variability presented could be from 3 samples collected within one month from two villages, from 300 samples collected across 12 villages, or from a large number of other possibilities. Without presenting this information, it is hard to know if deep-well tap water and not another transmission route was responsible for the reduction in enteric infectious diseases.

Wright et al 2004

One systematic review has been conducted to examine the variability of source and point of use water quality. Only 15 out of 57 studies measured *E. coli*; 11 remaining studies measured total coliforms and 22 studies measured fecal coliforms (Wright, Gundry et al. 2004). In half of the 57 studies, point of use water quality had higher contamination levels than source water quality (Wright, Gundry et al. 2004).

Given the differences in population density, remoteness, and variety of water source choice in rural areas that may lead to differences in water quality, it was surprising that Wright did not stratify by rural and urban studies. The sample size for each category was too small. Even so, there is considerable water quality variation across studies.

Variation between different studies using *E. coli* could not be explained by the study design, quality, or setting; however these variables were explanatory when predicting total coliform and fecal coliform counts. Of the six rural studies measuring *E. coli*, only four were published (Wright, Gundry et al. 2004). Three out of five studies demonstrated that water quality was worse at the household level than at the source (Morin et al 1990, Pinfold et al 1990, and Wright's unpublished study) while two studies had mixed results (El Attar et al 1992 and Simango et al 1992) (Wright, Gundry et al. 2004).

Water quality did not improve statistically over time between the collection source and the point of use demonstrating that recontamination is a reliably widespread and consistent issue for point of use strategies (Wright, Gundry et al. 2004). Recontamination after source improvement and household water treatment is a critical documented issue that household water treatment and safe storage seeks to address (Gasana, Morin et al. 2002; Trevett, Carter et al. 2004; Wright, Gundry et al. 2004; Fewtrell, Kaufmann et al. 2005; Kasirye 2010).

(Musa, Shears et al. 1999)

In May and October of 1996, 54 water samples were collected from a rural population living along the Nile River. 10 samples were taken from nomads living in the same areas. The geometric means of fecal coliforms was 2.7 times higher for the Nile River 2.5 times higher for the canal, and 3.7 times higher for piped water samples in October during the end of the rainy season than in May during the end of the dry season (Musa,

Shears et al. 1999). Households had better water quality than source water quality. Musa et al acknowledge the difficulty in associating diarrheal prevalence at the district level with local water quality values.

Household sampling methodology and selection were not presented. This may bias the direction of household water quality towards higher or lower levels of contamination depending on the household. It is unclear as to whether the samples were taken during the same day, one week, or over all weeks in May and in October.

(Molbak, Hojlyng et al. 1989)

260 households from three villages in rural Liberia were included in the study of stored water quality and diarrheal cases. Samples were taken from the households' stored water, source water, and small amounts of food were sampled for coliforms using the H_2S method. Out of the 20 source samples, half had levels of contamination of enterobacteria in the 10^4 to 10^5 log levels. A greater proportion of the household samples had higher log levels compared to the source water; 20 percent of the household samples had 10^5 to 10^6 log levels compared to 5 percent in the source water group (Molbak, Hojlyng et al. 1989).

(El Attar, Gawad et al. 1982)

Water samples from 10 household taps and 107 zirs, porous household storage containers, in Abbis II village, Egypt were collected and analyzed using the mTec method for *E. coli* and m-E method for measuring enterococci (El Attar, Gawad et al.

1982). Water was pretreated before being supplied to the villagers by the local government. *E. coli* and enterococci counts were higher in zir household water than in tap water. No significant changes in counts were observed after storage or between the summer and winter season (El Attar, Gawad et al. 1982).

Sampling collection of households' water was not clearly described. If the zir water came from household taps, then the lower levels of contamination in the household taps and higher contamination in zirs may indicate bacterial growth and contamination during storage. Yet, since only 10 taps were measured for water quality, there may be greater variability within the village taps.

(Morin, Jost et al. 1990)

In Butare, Rwanda, water quality of 100 springs was tested at four sites throughout the process of capping these springs at baseline. Next, springs and stored household water was tested daily during phase two. Rainfall was correlated with spring pollution. During the rainy season, springs had higher levels of contamination. Some families received Katadyn filters that reduced their total coliform counts and *E. coli* counts from 62% to 8% contamination (Morin, Jost et al. 1990). Protective fencing at the spring and gravel-sand filters did not significantly reduce the load of microorganisms (Morin, Jost et al. 1990).

The sampling methodology was not specified; therefore, it is difficult to determine the internal and external validity of the study.

(Pinfold 1990)

10 households' stored water quality and hand hygiene was tested five times for *E. coli* and salmonella contamination over eight months in northeastern Thailand. The fingertips of a household member were rinsed and tested for fecal coliforms and faecal streptococci. At the source, tubewell water had the highest quality; yet, households that used tubewell showed the highest levels of *E. coli* contamination (Pinfold 1990). Stored water used for drinking and cooking was shown to be cleaner than water used for washing dishes, bathing and using the toilet ($p<0.0001$) (Pinfold 1990). This study was very small and results are not necessarily representative of the area.

(Simango, Dindiwe et al. 1992)

In Mazowe, Zimbabwe, 216 water samples were collected from household storage containers of farm workers and 43 samples were collected from public standpipes. These samples were analyzed first for *E. coli* and then, if positive, 10 colonies from each test were cultured for pathogenic strains of *E. coli.* Water samples were cultured for salmonella, campylobacter, shigella, aeromonas, and *yersinia enterocolitica*. Food was also collected and analyzed in the same way for all pathogens.

Aeromonas was isolated from a higher percentage of stored drinking water than from food. Enterotoxigenic *E. coli* and *campylobacter* were isolated from stored drinking water. The majority of household stored drinking water had *E. coli*; yet, most household drinking water was underneath 20 *E. coli* per 100 milliliters (Simango, Dindiwe et al. 1992). Water storage containers had higher levels of *E. coli* when compared to water

collection containers that was similar to El Attar's findings (Simango, Dindiwe et al. 1992). Simango and colleagues did not identify how households were selected or give a sampling time frame that are necessary for determining study validity and bias.

(van der Hoek, Konradsen et al. 2001)

In Hakra, Pakistan, 200 households were randomly selected from 10 randomly selected villages. All sources used by the 200 households and a random selection of 50 households were sampled weekly for water quality and analyzed for *E. coli* using the m-coli Blue 24 method (van der Hoek, Konradsen et al. 2001). Water from irrigated seepage sources was generally of higher quality than surface water (49% compared to 22%)(van der Hoek, Konradsen et al. 2001). When water quality was controlled for in the logistic regression models, water quantity was significant in predicting diarrheal risk ratios.

(Blum, Emeh et al. 1990)

A project where communities were given boreholes with hand pumps, ventilated improved pit latrines, and hygiene education was piloted in Imo State, Nigeria. The project aimed to decrease diarrhea and guinea worm transmission. Once during the dry season and once during the wet season enumerators took water samples from 12 control and 12 intervention households at the household's water source, from the carrying container, and after household storage (between 2 and 24 hours after collection)(Blum, Emeh et al. 1990). Water samples were analyzed for fecal coliforms and streptococci counts per 100 milliliters of water. Carrying containers and storage

containers of borehole users showed contamination levels 10^2 to 10^4 per 100ml higher than source contamination (0 to 27 fecal-colony units per 100 milliliters) (Blum, Emeh et al. 1990).

The study lacks water quality baseline information. It is unclear when the water samples were taken during the study and if the same 12 households in each group were observed in the dry and rainy seasons.

(Heinanen, Chandiwana et al. 1988)

Farmers' main drinking water quality was studied in the Burma valley, Zimbabwe. Samples of water were collected from the two main water sources of each of the 20 commercial farms and analyzed for fecal coliforms using membrane filtration. Samples were collected three times in the first week of February, March, and April between 7:30 am and 5:30 pm. In March, household water quality of many containers was taken from three compounds.

Across the study, boreholes had higher water quality than river water, piped river water, piped mountain water, piped dam water, and well water. Piped mountain water had higher quality in March than in February; besides this, the mean fecal coliform counts for each source did not show seasonal differences (Heinanen, Chandiwana et al. 1988). Water temperature was higher than 15 degrees Celsius and the pH ranged from five to eight.

(Lehmsuluoto 1986)

In Western Kenya, water sources and household users' stored water quality were studied to provide guidance on whether or not these should be included in impact evaluations of interventions. Water samples were collected once from 78 unprotected springs, 33 protected springs, 6 shallow wells, 2 boreholes, and 216 household users. Samples were analyzed for fecal coliforms. 54% of unprotected springs, 50% of the protected springs, and 78% of households had 10 or more fecal coliforms per 100 ml (Lehmsuluoto 1986). Recontamination was present for household users of every source with households using protected springs and boreholes showing the least levels of contamination (68% and 71% exceeded 10 fecal coliforms per 100 ml) (Lehmsuluoto 1986). Lehmsuluoto and colleagues concluded that studies should measure water quality routinely as an indication of progress.

Selection of household users and number of households per source was not described. It could be assumed that 1 to 3 households per source were sampled, but this is unknown. This could have resulted in selection bias as households' water nearest the source may have been sampled in an effort to save time.

(Platenburg and Zaki 1993)

In 20 villages in Upper Egypt, hand pumps and latrines were installed in 10 villages to examine if this intervention would reduce childhood diarrhea. In each village, hand pumps and latrines were installed for every 8 to 10 households. Water quality was taken from 10 selected hand pumps in each of the control villages and from 10 selected

sources used in the intervention villages; household water quality was also sampled from all households that used these selected sources (Platenburg and Zaki 1993). Hand pump water was tested for fecal coliforms using the multiple tube method with 5 tubes at installation and after 3, 6, and 11 months. With this testing method, the detectable range of most probable number of *E. coli* was small (<2.2MPN to >16.0MPN). At installation, 90 percent of the hand pumps showed contamination greater than 2.2 MPN per 100ml; water quality improved at the source over time (Platenburg and Zaki 1993).

There was no significant difference between the contamination of intervention and control households as all had the highest levels of contamination throughout the study; this means significant household recontamination was exhibited after collection (Platenburg and Zaki 1993). It is possible that the 5 tube testing method was not sensitive enough to detect changes in the household water quality. If IDEXX trays or membrane filtration with a wider detection range were used, then perhaps changes in the intervention's water quality could have been detected.

(Rajasekaran, Dutt et al. 1977)

In five villages of Athoor block, India, the impact of water supply on diarrhea and shigellosis was studied. Water samples were collected each month from source samples: 8 wells, 10 street taps, and 8 household taps. Twice a month, 70 samples were collected representative of source usage from households reflecting 10 percent of the larger population. All samples were analyzed for shigella although it was isolated in drinking water primarily during the monsoon season (October through December). 86

to 88 percent of all household water storage samples had levels above 10 MPN of *E. coli* per 100 milliliters (Rajasekaran, Dutt et al. 1977). Although water was collected on a monthly basis for 11 months, seasonality of water quality was studied. Given a large sample of 116 samples each month, future analysis could provide useful information.

(Shears, Hussein et al. 1995)

In Matipara, Rajbari, and Selimpur, Bangladesh, microbial resistance was assessed using samples from source waters, household water, and stool samples from children's feces. Feces samples were exposed to five antibiotics and sensitivity analyses were conducted. Water from tube wells, storage ponds, rivers, streams, and household pots was analyzed using membrane filtration method. Tube well water was of the highest quality (<10 colony-forming units per100ml); 48% of household stored water was above 50 colony-forming units per 100 milliliters (Shears, Hussein et al. 1995). 62% of households using high quality tube well water had evidence of recontamination after water collection. Multiply resistant enteric bacteria were discovered to be widely present in the water sources, 76% of household storage containers, and in the children's feces (Shears, Hussein et al. 1995).

(Sutton and Mubiana 1989)

In 10 selected villages in western Zambia, water quality was taken from 25 randomly selected households and analyzed for total coliforms and for fecal coliforms. Villages were not selected into the study that had source water quality with more than 10 coliforms per 100 ml. Of the 85 percent of household water samples that were below

the detectable limit, 58 percent of households used traditional sources and 42 used improved sources (Sutton and Mubiana 1989). Population density in a village or town was reported to be a high risk factor for greater household contamination (Sutton and Mubiana 1989).

(Trevett, Carter et al. 2004)

In three rural Honduran villages, household stored drinking water was collected four times from 43 households over two years. 60% of the households used hand-dug wells whereas 40% used boreholes. All household stored drinking water quality samples showed contamination between the source and the household. Households using boreholes had lower levels of thermotolerant coliform contamination than those using hand dug wells, but they also had a higher proportion of recontamination when compared with the source water quality. Authors found that household water quality varied widely on any given day although they did not state the level of variability. Time stored, seasonality, and rainfall had no effect on stored water quality. It is possible that these effects were not seen because household stored water quality was either categorized using the WHO's drinking water categories or the geometric mean of all households drawing water from one source was used as the unit of analysis.

The methodology of selecting households is weak as a convenient sample was collected that undermined the internal and external validity of the study. These households could have been selected into the study because they had particularly poor water collection

behaviors. It is also not documented how often samples were taken which limits the study's broader applicability.

(Verweij, van Egmond et al. 1991)

In four villages in Venda, South Africa, water samples were collected from two chlorinated taps, five boreholes, and nine unprotected springs. In each village, one to three households' water was taken immediately after filling the storage container, two and four hours after storage, and after overnight storage. Samples were analyzed using the membrane filtration method. Water quality varied by source; tap water had no detectable fecal coliforms in the samples. Water quality information collected from two hours, four hours, and overnight storage was not presented.

(Empereur-Bissonnet, Salzman et al. 1992)

In Boassa village in Burkina Faso, water quality and storage practices were studied to determine the transportation pollution levels of water from the source to the household. Each of the 29 households in the village was surveyed on water storage practices; water samples were taken before households were given a new metal water storage container. Water samples were taken directly from the pump, right after a household returned from collecting the water in the original collection container, at the household using the household's storage container, and 24 hours after storing the collected water. Water samples were analyzed for total and fecal coliforms using the membrane filtration method. No fecal or total coliforms were detected at the source pumps. 38% of the households contaminated their water in transporting it back to their

households: 131 total coliforms and 96 fecal coliforms were detected on average per 100 milliliter samples (Empereur-Bissonnet, Salzman et al. 1992). New water storage containers and education improved water quality during transport from 96 to 11 fecal coliforms per 100 milliliters, at storage from 207 to 59 fecal coliforms per 100 milliliters, and of drinking water from 349 to 69 fecal coliforms per 100 milliliters (Empereur-Bissonnet, Salzman et al. 1992). Households did not reach the WHO's recommended levels of safe water.

(Shiffman, Schneider et al. 1978)

The link between water quality and food wastage was studied over three years in two rural villages: Florida Aceituno and Guanagazapa, Guatemala. An advanced treatment and distribution system was installed which included chlorination. In Guanagazapa, 97% of water samples collected from the distribution system did not test positive for coliforms and 65% of household samples tested positive for coliforms (Shiffman, Schneider et al. 1978). In Florida Aceituno, 59% of shallow well samples and 52% of households' stored water samples tested positive for coliforms (Shiffman, Schneider et al. 1978). Shiffman and colleagues did not detect a significant decrease in food wastage or in decreased diarrheal cases over the three years.

Sample size, selection of households, time at which water samples were taken, and frequency of testing were not reported which are necessary for determining the internal and external validity of the study.

(Feachem, Burns et al. 1978)

58 villages from the Mafeteng, Maseru, and Mokhotlong districts were representatively selected. 703 households were studied from one to four days. 43% of the household samples came from the Mafeteng district located in the lowlands. Water samples were taken at the households' preferred source and stored water from all households using a source was tested for *E. coli* and for fecal streptococci using the membrane filtration method.

Most water samples were processed within eight hours after sampling but some were processed within 12 hours after sampling. A small study was conducted to see how sampling time influenced the results comparing the growth or die-off at 12 hours to 8 hours after sampling. The confidence interval for fecal coliforms was between a 10% increase and decrease of 9%; the confidence interval for fecal streptococci was between a 3% increase and 28% decrease at 12 hours after sampling (Feachem, Burns et al. 1978). Authors conclude that this did not significantly bias their results.

For 69% of unimproved sources, the sources had five times higher log concentrations of fecal coliforms in the wet season than in the dry season(Feachem, Burns et al. 1978). From the source to the households stored water containers, fecal coliforms increased by 10^3 coliforms per 100 milliliters (Feachem, Burns et al. 1978). Feachem and colleagues suspected that diarrhea and typhoid in the wet season were not due to water-borne transmission; however, they acknowledged that their monitoring strategy might not have been intensive enough to detect the differences.

(Kaltenthaler, Drasar et al. 1996)

In northern Botswana, samples were collected from 110 households in two villages. Caretakers were asked to dip the cup usually used for dipping into the water container to provide water for the study (Kaltenthaler, Drasar et al. 1996). Water samples were analyzed using the membrane filtration method. Source water had different levels of contamination: the river (0 cfu), private standpipes (mean 0.25 cfu), public standpipes in village A (mean 0.5 cfu), and stream (10 cfu) water had the lowest counts per 100 milliliters compared to the two ponds (220 cfu and 1,000 cfu) and the public standpipes in village B (mean of 112 cfu) (Kaltenthaler, Drasar et al. 1996). Water samples from each village had no statistically significant difference in contamination; the mean of village A was 53 colony-forming units per 100 milliliters whereas the mean for village B was 95 colony-forming units per 100ml. This study also measured contamination of plates, fingers, and cloths.

(Lindskog and Lindskog 1988)

Near Zomba Mountain in Malawi, water samples were collected from springs, wells, rivers, pipes, and stored household water during the beginning and peak of the dry season and during two rainy seasons. Household samples represented the proportion of users of the source and were analyzed for fecal coliforms, total coliforms and streptococci using the membrane filtration method (Lindskog and Lindskog 1988). Seasonality of water sources was studied.

Water quality was worse during the rainy season than during the dry season for wells, springs, and piped water. Household water quality and source water quality were statistically significantly correlated ($p<0.01$) (Lindskog and Lindskog 1988). Households' source switching to less contaminated sources during the rainy season matched the microbiological contamination patterns (Lindskog and Lindskog 1988).

(Mazengia, Chidavaenzi et al. 2002)

In rural Zimbabwe, 30 intervention households were provided with new water urns and 30 households were controls. All households used upgraded protected wells. Two 250-milliliter samples were collected from each well four weeks apart; water was sampled at each time from the household containers or urns. When comparing water urns to well supply, the water urns had a protective effect on water quality ($p<0.01$) (Mazengia, Chidavaenzi et al. 2002). The normal water storage containers had similar water quality of supply wells indicating potential contamination ($p>0.05$) (Mazengia, Chidavaenzi et al. 2002). Yet, Mazengia and colleagues note that households with water urns had higher well contamination levels that may have accounted for the detectable difference in the intervention. This indicates that water quality may have been different for households using regular storage containers if they had higher source water contamination levels.

(Cidavaenzi, Jere et al. 1998)

In Zimbabwe, 60 households in Rota and Chiviya villages were selected for a controlled trial of water urns. Half of the households were given new water urns. In a four-week

period, water samples were collected twice between 7 am and noon from the urns and from controls' household water storage containers. These samples were analyzed for fecal and total coliforms using the membrane filtration method. After collection, water from the urns improved significantly from the supply well water ($t=3.97$, $p<0.01$). Household water in the control group was observed to be the same as well water ($t=0.2$, $p>0.05$). Household water quality of urn users was significantly higher than traditional household storage container users ($t=2.88$, $p<0.01$).

(Esrey, Habicht et al. 1986)

In Lesotho, source and household water was tested for fecal coliforms and for fecal streptococci and analyzed via the membrane filtration method. 10 villages were selected as controls and in 10 villages water quality was improved. In each village, 294 children were followed over three phases each lasting five weeks. Stool samples from children was collected from the households and analyzed for *Giardia lamblia*, *E. coli,* and *Campylobacter*. Only 44 percent of children provided stool samples. Children whose households used source water quality below 10 fecal streptococci per 100 milliliters had a lower infection rate of *giardia, E. coli,* and *campylobacter* than children of households that used poorer source water quality (Esrey, Habicht et al. 1986). Diarrhea rates and child growth rates of households that used better sources were not significantly better than those households using poorer source water quality. Older children showed more weight gains than younger children when fecal *streptococci* counts were low (Esrey, Habicht et al. 1986). The study time is very short: three study months over nine months. It is not explained in the study as to why nine months was

thought to be a long enough period to document improved and significant weight gains or growth rates.

(Sandiford, Gorter et al. 1989)

150 water samples were collected from Villa Carlos Fonseca, Nicaragua across wet and dry seasons on a monthly basis to determine the drivers of water quality. Household and source collection sites were identified based on a population survey stratified by communities. Samples were analyzed for *E. coli* within eight hours of collection; this is above the WHO's recommended six hours waiting time and could have biased the results in either direction as bacteria grow and die after collection. Samples were disproportionately taken during the dry season; 20 samples were taken during the wet season and 33 samples were taken during the dry season(Sandiford, Gorter et al. 1989). Water quality determinants were the type of water source, season, community size, storage practices, and interaction between source type and rainfall. Rivers, streams (10^6 - 10^{12}), and protected bucket wells (10^1-10^{11}) had higher levels of log fecal coliforms in the dry season than unprotected wells and springs (10^2-10^9), protected wells with pumps (10^0-10^6), public standpipes (10^0-10^5), and household connections (10^0-10^2)(Sandiford, Gorter et al. 1989). Sandiford and colleagues hypothesize that protected domestic wells may have greater fecal contamination than unprotected common wells and springs due to their proximity to animals, households, and children who openly defecate.

(Tomkins and Drasar 1978)

Tomkins and colleagues studied the association between water quality and anthropometric indices of children ages 3 to 49 months in Gamzago, Nigeria. Household water was collected and analyzed for fecal coliforms. Anthropometrics were not clearly linked with protected or unprotected water sources; fecal coliforms were higher from unprotected water sources than from protected and stored sources(Tomkins and Drasar 1978). Water sample size and collection methodology was not clearly defined in the study.

(Levy, Hubbard et al. 2009)

Levy and colleagues studied the drivers of water quality variability in a rural village of Colon Eloy in Northern coastal Ecuador. Since 72% of the villagers used the stream as their primary drinking water source, sampling centered on being able to describe the hourly, daily, weekly, and monthly water quality variability of the stream. Each week, seven randomly selected households' stored water was collected. In contrast to Trevett and other studies, household water samples in both dry and wet seasons were 0.7 log *E. coli* levels below source water quality levels. In the dry season, these results may have been due to a greater proportion of water household water treatment and to water collection from other sources. The geometric mean of household water quality and stream water quality was only loosely correlated over time (r^2=0.15, p=0.006) (Levy, Hubbard et al. 2009). This correlation may have been stronger if the same households were sampled each week allowing for a time series comparison. However, only half of the households were even asked whether they were going to use the water for drinking

or not as the purpose of the collected water was not seen to drive household water quality. It is not surprising then, that a stronger seasonality association was observed in source samples than in household samples. Stream water quality in the dry season had lower *E. coli* counts but higher hourly (2.35 log differences), daily (1.0-2.2 log differences), and weekly variability (1.8 log difference) (Levy, Hubbard et al. 2009). In the wet season, stream water quality showed 0.4 log levels higher of *E. coli* than in the dry season (Levy, Hubbard et al. 2009).

This study is important for describing the water quality variability at the source, but it does not sufficiently examine stored household water quality. Since source water quality varies more based on hourly collection basis than on daily or weekly basis, Levy and colleagues strongly recommend taking more than one grab sample when testing water quality. Since the stream was found to be least contaminated in the morning before people started bathing and washing clothes in the river, it also follows that it is best to take water samples when villagers actually collect their drinking water.

(Levy, Nelson et al. 2008)

Household and stored water was collected from five villages in northern coastal Ecuador. If a household used rainwater or treated drinking water, 50 milliliters was tested whereas if a household used surface or well water, then 10 milliliters was tested using a modified membrane filtration test for *E. coli* and for *Enterococci*. In the first visit of this study, 39 households with reported diarrhea and 82 control households without reported diarrhea were selected to participate. Household water samples were taken

from three different household water storage containers; only three containers were sampled if households had more than three containers. Household samples were then taken to a controlled field lab and stored with closed lids. Each day that the household still used the drinking water, samples were taken from the household's water stored at the field lab and directly from the household to measure the microbial growth and die-off and household recontamination.

The type of container and water treatment was significant for the presence of *Enterococci* (container p-value=0.001, treatment p-value=0.05) but not for *E. coli* (container p-value=0.68, treatment p-value=0.25) (Levy, Nelson et al. 2008). The type of water source was only significant for predicting log levels of *E. coli* (Levy, Nelson et al. 2008). *E. coli* and *Enterococci* levels of stored drinking water decreased through the third day of storage when compared to controlled household stored drinking water kept in the field lab (Levy, Nelson et al. 2008). After the third day, evidence of recontamination was present in half of the households (Levy, Nelson et al. 2008). This study is unique because it is the first to compare field-lab control household drinking water quality to household water quality in real time. Levy et al report that this careful study design partially explains why source samples had higher observed microbial loads than observed household samples. Past studies of household recontamination have used sources with high initial water quality; yet Verweij et al and Musa et al also observed better household water quality than source water quality (Levy, Nelson et al. 2008).